ISSN 2280-7853

PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

FIRENZE UNIVERSITY PRESS

Topics
Phenomenology and Social Ontology; Ethics and Political Theory; Cognitive Neurosciences, Philosophy of Mind and Language, Logic; Aesthetics, Metaphysics and History of Ideas.

Frequency
2 issues per year

Editor-in Chief
Roberta De Monticelli (PERSONA)

Co-Editors
Research Centers
Roberta Sala (CeSE P)
Matteo Motterlini (CRESA)
Andrea Tagliapietra (CRISI)

Faculty
Claudia Bianchi, Massimo Cacciari, Massimo Donà, Roberto Mordacci, Massimo Reichlin

Vice-Editor
Stefano Cardini

Managing Editor
Francesca De Vecchi

Editorial Team
Stefano Bacin, Francesca Boccuni, Emanuele Bottazzi, Emanuele Caminada, Francesca De Vecchi, Francesca Forlé, Diego Fusaro, Alfredo Gatto, Giuseppe Girgenti, Roberta Lucentini, Barbara Malvestiti, Francesca Pongiglione, Andrea Sereni, Elisabetta Sacchi, Sarah Songhorian, Marco Spina, Silvia Tossut, Francesco Valagussa

Graphic Design
Dondina e associati (print version)

Graphic Layout
Direweb (on line version)

Web Site Editorial Board
Emanuele Caminada, Stefano Cardini, Francesca Forlé, Barbara Malvestiti, Sarah Songhorian, Marco Spina

Registrazione del Tribunale di Pavia, n° 6 del 9/07/2012
Direttore responsabile: Roberta De Monticelli
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

n. 11 - 2016

PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

EMOTIONS, NORMATIVITY, AND SOCIAL LIFE

Edited by Francesca Forlè and Sarah Songhorian

FIRENZE UNIVERSITY PRESS

Phenomenology and Mind practices double blind refereeing and publishes in English.

CC 2016 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
www.fupress.com

Phenomenology and Mind. The Online Journal of the Faculty of Philosphy, San Raffaele University, on-line: http://www.fupress.net/index.php/pam

Stefano Canali (Scuola Internazionale Superiore di Studi Avanzati - SISSA)
Ian Carter (Università degli studi di Pavia)
Emanuela Ceva (Università degli studi di Pavia)
Antonio Da Re (Università degli studi di Padova)
Mario De Caro (Università di Roma III)
Corrado Del Bo (Università degli studi di Milano)
Emilio D'Orazio (POLITEIA - Centro per la ricerca e la formazione in politica ed etica, di Milano)
Maurizio Ferrera (Università degli studi Milano)
Luca Fonnesu (Università degli studi di Pavia)
Anna Elisabetta Galeotti (Università del Piemonte Orientale, Vercelli)
Barbara Herman (University of California, Los Angeles - UCLA)
John Horton (Keele University)
Andrea Lavazza (Centro Universitario Internazionale di Arezzo)
Eugenio Lecaldano (Università degli Studi di Roma "La Sapienza")
Neil Levy (University of Melbourne)
Beatrice Magni (Università degli studi di Milano)
Filippo Magni (Università degli studi di Pavia)
Massimo Marassi (Università Cattolica di Milano)
Alberto Martinelli (Università degli Studi di Milano)
Susan Mendus (University of York)
Glyn Morgan (Syracuse University in New York)
Anna Ogliari (Università Vita-Salute San Raffaele)
Valeria Ottonelli (Università degli studi di Genova)
Federico Gustavo Pizzetti (Università degli Studi di Milano)
Mario Ricciardi (Università degli studi di Milano)
Nicola Riva (Università degli Studi di Milano)
Adina Roskies (Dartmouth College)
Giuseppe Sartori (Università degli Studi di Padova)
Karsten R. Stueber (College of the Holy Cross)
Nadia Urbinati (Columbia University)
Corrado Viafora (Università degli studi di Padova)

Cognitive Neurosciences, Philosophy of Mind and Language, Logic (CRESA)
Edoardo Boncinelli (Università Vita-Salute San Raffaele)
Stefano Cappa (Institute for Advanced Study, IUSS, Pavia)
Benedetto de Martino (University College London, UCL)
Claudio de' Sperati (Università Vita-Salute San Raffaele)
Michele Di Francesco (Institute for Advanced Study, IUSS, Pavia)
Massimo Egidi (Libera Università Internazionale degli Studi Sociali Guido Carli di Roma, LUISS Guido Carli, Roma)
Francesco Guala (Università degli studi di Milano)
Vittorio Girotto (Istituto Universitario di Architettura di Venezia, IUAV, Venezia)
Niccolò Guicciardini (Università degli studi di Bergamo)
Diego Marconi (Università degli studi di Torino)
Gianvito Martino (Università Vita-Salute San Raffaele)
Cristina Meini (Università del Piemonte Orientale)
Martin Monti (University of California, Los Angeles, UCLA)
Andrea Moro (Institute for Advanced Study, IUSS, Pavia)
Michael Pauen (Berlin School of Mind and Brain, Humboldt-Universität)
Massimo Piattelli Palmarini (University of Arizona)
Giacomo Rizzolatti (Università degli studi di Parma)
Marco Santambrogio (Università degli studi di Parma)
Achille Varzi (Columbia University)
Nicla Vassallo (Università di Genova)

CONTENTS

CONTENTS

INTRODUCTION

INTRODUCTION

Francesca Forlè, Sarah Songhorian
The Emotions: From the (Inter)Personal to the Normative Dimension

FRANCESCA FORLÈ
Vita-Salute San Raffaele University
francescaforle@hotmail.it

SARAH SONGHORIAN
Vita-Salute San Raffaele University
songhorian.sarah@unisr.it

THE EMOTIONS: FROM THE (INTER) PERSONAL TO THE NORMATIVE DIMENSION

abstract

Recent debate in the intersection between philosophy and cognitive sciences have underlined the relevance of emotions in our lives. Emotions seem to play several different roles: from our inner and personal experience of the world, especially in all its alleged value aspects, to our relations to others. Yet, philosophical and scientific debate has not reached a consensus on what emotions are and on their specific contribution to our personal and social lives.
In this introductory chapter we present the different debates on emotions this issue focuses on and we briefly summarize the papers collected here.

keywords

Spring School of Philosophy 2016, theories of emotions, emotions morality and political theory, hate speech, emotions and intersubjectivity

Phenomenology and Mind, n. 11 - 2016, pp. 10-17
DOI: 10.13128/Phe_Mi-20102
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

San Raffaele Spring School of Philosophy 2016

The present issue of *Phenomenology and Mind* "Emotions, Normativity, and Social Life" contains a selection of papers presented at San Raffaele Spring School of Philosophy (SRSSP) 2016 both by invited speakers and by contributors selected through double-blind peer review. Moreover, it includes papers that were selected for presentation but whose authors could not attend SRSSP 2016.
SRSSP was held at Vita-Salute San Raffaele University in Milan in June, 6th – 8th, 2016. It was organized by the research centers CeSEP, CRESA, and PERSONA of the Faculty of Philosophy and it also received support from the Faculty and from the Ph.D. program.

Emotions and the Affective Life

Recent debate in the intersection between philosophy and cognitive sciences have underlined the relevance of emotions in our lives. Emotions seem to play several different roles: from our inner and personal experience of the world, especially in all its alleged value aspects, to our relations to others. As for the former, the way we perceive the world and understand it is deeply influenced by our emotions. As for the latter, several researches have shown that emotional engagement with others is crucial both for understanding them and their states of mind and for building a relation with them. Just to provide a few paradigmatic examples of this interpersonal role, one can think of the constitution and maintenance of the bond between a newborn and his or her primary care-giver as an emotional one. Similarly, research on moral dilemmas have underlined that emotions play a crucial role in what we decide to do or to avoid doing (Greene *et al.* 2001). Yet, besides these roles, philosophical and scientific debate has not reached a consensus on what emotions are and on their specific contribution to our lives. There is great amount of data on the fact that they play a role, but what exactly that role is or what relation emotions entertain with other human faculties and abilities is not yet clear. Furthermore, the epistemological status of emotions is controversial.
Moreover, theories of emotions constitute a relevant attempt to bring empirical work to bear on philosophical theorizing, so providing the touchstone for interdisciplinary analysis.
Some of the more interesting debates concerning emotions focus on their role in moral and political theory, on their relation to language and hate speech, on their role in intersubjectivity – both in typical development and in pathologies –, and on our emotional self-understanding. Focus on each of these dimensions provides a more detailed understanding of what emotions actually are and of how they guide our lives.

Four Major Debates in the Field

Within this framework, papers have been selected to contribute to one of the following sections:

– *Theories of Emotions*

What are emotions? What relationships do they entertain with perception, thought, and rationality on the one hand, and with their bodily and social expressions on the other? The aim of this section is to focus on these topics, trying to contribute to the contemporary debate on the ontological and epistemological status of emotions (Frijda 1986, 1988; Goldie 2000; Nussbaum 2001; Prinz 2004; Whiting 2011) and on their role in our personal and social life (Gallagher 2001, Zahavi 2014).

– *Emotions, Morality, and Political Theory*

The debate in morality and in political theory on whether emotions do and should play a role in our decision-making has been crucial for centuries. New life has been given to this debate by the general renaissance of emotional theories linked in particular to psychological and social research. The questions guiding this section are: What emotional constructs or abilities – if any – play a role in our moral and political understanding of ourselves, of others, and of the world surrounding us? What consequences can derive from an emotional understanding of morality and political theory – that is, are emotions a good candidate for explaining them? What specific emotions – if any – constitute a touchstone for political and moral decision-making? Are emotions – or should they be – the only mechanism to pass moral and political judgment or is there or should there be something else too?

– *Emotions, Language, and Hate Speech*

The debate on hate speech and slurs underlines how language can convey emotions and phobias as well as bring about negative emotions on the target of derogatory language. Moreover, it focuses on the evaluative properties of slurs and derogatives. The guiding questions for this section are: What are slurs and hate speech? What kind of evaluative properties – if any – do slurs have? What kind of emotional, psychological, and societal consequences do slurs and hate speech have? Should we regulate their usage by limiting freedom of speech?

– *Emotions and Intersubjectivity – Typical Development and Pathologies*

Developmental research has pointed to the emotional environment in which a child grows up as an indicator of how capable he or she would later be to enter in relation to others and to understand them (Kohut 1977; Stern 1990; 1985; Hofer 2006; 1981; Schore 1994; Bornemark 2003). Emotions and intersubjectivity seem to be strongly linked from a developmental perspective. Thus, on the one hand, this section elaborates on this relation. On the other hand, this section focuses also on the relationship between emotions, emotional regulation, and pathologies – such as, Borderline Personality Disorder (BPD) and Autism Spectrum Disorder (ASD).

Contents

Dieter Lohmar (Husserl-Archiv der Universität zu Köln) opens the section *Theories of emotions* with his invited paper "Emotions as a Multi-modal System of Representation in Non-linguistic Thinking". In this paper, Lohmar claims that emotions can carry meaning, evaluations, knowledge, experience and sometimes also plans for the future. For this reason, he believes they can function as part of non-linguistic thinking. He thus investigates the function of emotions as an important part of an effective system of thinking and deciding that does not use the concepts of language but visual scenes combined with feelings.

In her invited paper, "Is the Perceptual Model of Emotions Still A Good Competitor? A Small Phenomenology of Feeling", Roberta De Monticelli (Vita-Salute San Raffaele University, Milan) defends a version of the Perceptual Model of feeling, according to which feeling is the mode of presence of axiological aspects of reality, or values. After considering the objections to

this Model, she presents an intentional analysis of feeling as the core of all emotions, more generally of all the phenomena of affective life, for which she proposes a taxonomy. She also presents a special phenomenology of the consciousness that feels and of the degrees of intuitive cognition of values which may be distinguished in it. Finally, she reconsiders the objections to the Perceptual Model in order to make sense of them, showing the extent to which they are valid, and the way to overcome them.
In "The Rationalities of Emotion", Cecilea Mun (Independent Scholar) argues that emotions are rational in-themselves and that they are also instrumentally rational, epistemically rational, and evaluatively rational. After discussing what it means for emotions to be rational or irrational in-themselves, she proposes a criterion for the ontological rationality of emotions. Finally, she discusses some of the implications of this account.
In his paper "Emotional Intelligence as an Intellectual Virtue: Theoretical Analysis and Empirical Assessment", Paul Poenicke (University of Buffalo) claims that virtue theory has long recognized the significance of emotion for cognition, yet little philosophical research has been dedicated to identifying an intellectual virtue related to emotion. According to the author, applying recent work in virtue epistemology reveals emotional intelligence (EI) to be an intellectual virtue. People who score high in EI better attend to epistemically-significant features of the environment, which could explain the significance of stakes for knowledge attribution. While testing did not confirm higher EI with stakes sensitivity (the hypothesis), study methods, including stakes vignettes, inadvertently caused the hypothesis to be disconfirmed.
The fifth paper of this section "Affectivity and Self-Displacement in Stein's Early Phenomenology. On the Role of Self-Experience in Empathy" by Elisa Magrì (University College Dublin) focuses on the role of bodily self-displacement in Stein's account of empathy, pointing out its relevance in the general dimension of affectivity. According to the author, Stein grounds empathy on a dynamic model of embodied self-experience, which shares significant similarities with Varela & Depraz's neurophenomenology. However, Magrì argues that Stein's view of empathy cannot be reduced to a naturalized phenomenological sense and that bodily self-displacement is pre-condition of a more complex disposition towards others as is in line with Ratcliffe's theory of radical empathy.
Sandy Berkovski (Bilkent University) closes the section with his paper "A naturalist View of Humiliation". The author argues that a naturalist analysis of humiliation begins with the notion of social interaction, a public encounter with other people. Interactions are an essential element in cooperation, a vital condition of survival and well-being. In the course of interaction, the person presents himself or herself as someone possessing the qualities necessary for successful cooperation. An act of humiliation is designed to inflict damage on the agent's self-presentation. Any such damage would be a sign that the agent is not successful in conducting the given interaction. Such damage would tend to decrease of cooperative value of the humiliated individual and to decrease his or her chances of survival and reproduction.

The invited paper "Straight and Twisted Self-deception" by Anna Elisabetta Galeotti (University of Eastern Piedmont "Amedeo Avogadro") opens the section *Emotions, Morality, and Political Theory*. The paper analyzes two different forms of self-deception – straight and twisted ones – claiming that they are both actual cases of self-deception, even though the way to account for them is not a unitary one, as in Mele. The paper critically examines the claim that in twisted self-deception the motivational state of the subject is dominated by emotions. Galeotti, thus, proposes an alternative explanation of self-deception in which emotions – together with wishes – play a role in both types of self-deception.
In his invited paper "Moral Emotion, Autonomy and the 'Extended Mind'", Edward Harcourt

(Oxford University, Keble College Oxford) considers a 'micro' thesis on shame and guilt and a 'macro' thesis on self-regulation. According to the former thesis, shame and guilt are equally other-dependent. Moreover, because other-dependence in either emotion is not a mark of heteronomy, neither emotion is more characteristic of a well-functioning moral consciousness. As for the 'macro' thesis, the author argues that the other-dependent view of self-regulation usually ascribed to children can also be found in adults. So that, if one assumes that only the moral consciousness of typical adults can be well-functioning, then there is all the more reason to think that other-dependence and a well-functioning moral consciousness can go together. If other-dependence can generally be a characteristic of our self-regulatory mechanisms when they function well – the paper's 'macro' thesis –, then there is all the more reason to accept the paper's 'micro' thesis – that other-dependence can characterize the well-functioning of both shame and guilt. The conclusion is that heteronomy lies not in the fact of other-dependence but in the nature of the dependence.
In her paper "Caring about an Ethics of Care: A New Dimension", Maria Giovanna Bevilacqua (University of Italian Switzerland) wonders whether ethics of care could provide a useful tool to understand some of the contemporary issues in our societies. In order to do so, she focuses on some themes that ethics of care underlines – namely, the responsiveness of the moral subject for others' needs and the difference between concrete other and generalized other.
Gian Paolo Terravecchia's (University of Padua) paper – "Social Stances, Emotions and the Importance of Fear" – presents five main social stances: to refuse, to suffer, to accept, to assent, and to make something one's own. According to the author, they all depend on different types of relationship between an interior attitude and an exterior manifestation. The second main contribution of the paper consists in a discussion of fear and its relationships to social stances. Studying emotions helps to stress the similarities and the differences between social stances and emotions, and among social stances themselves. The paper gives an example of how ethics can be enlightened by the tools of social philosophy.

Robin Jeshion (University of Southern California) opens the section *Emotions, Language, and Hate Speech* with her invited paper "Slur Creation, Bigotry Formation: the Power of Expressivism". In this paper, Jeshion presents two novel problems to be added in the current debate on slurs. The *Slur Creation Problem*: How do terms come to be slurs? An expression 'e' is introduced into the language. What are the mechanisms by which 'e' comes to possess properties distinctive of slurs? The *Bigotry Formation Problem*: Speakers' uses of slurs are a prime mechanism of bigotry formation, not solely bigotry perpetuation. With a use of a slur, how are speakers able to introduce new bigoted attitudes and actions toward targets? She concludes by arguing that expressivism offers powerful resources to solve the problems.
In his paper "Slurs: Semantic Content, Expressive Content and Social Generics", Federico Cella (Vita-Salute San Raffaele University, Milan) discusses the thesis that the offensiveness of slurs is related to the semantic encoding of stereotypes in their meaning. As noticed by Robin Jeshion, the stereotypical strategies do not seem to provide a satisfactory analysis of slurs' functional traits. For this reason, Cella proposes to modify her view by making a distinction between two offensive dimensions of slurs: a negative expressive component encoded in the semantic content and directed toward a certain group of individuals, and the social generics related to that group conveyed as conversational implicatures.
In her paper "Not All Slurs are Equal", Mihaela Popa-Wyatt (LOGOS, University of Barcelona) argues that, in contrast with the standard view that slurs convey contempt based on group-membership, slurs are not a unitary group. In order to defend her thesis, she analyses two dimensions of variation among derogatives supporting the thesis that contempt based on group-membership doesn't cover all the data.

In the fourth paper of this section – "Building Evaluation into Language" –, Bianca Cepollaro (Scuola Normale Superiore, Pisa and Institut Jean Nicod, Paris) spells out the conditions for a uniform analysis of thick terms and slurs. The main claim is that thick terms and slurs convey evaluations via presupposition and represent a device through which language implicitly conveys linguistically encoded evaluations.
In their paper "Slurs and Negation", Francesca Panzeri (University of Milan-Bicocca) and Simone Carrus (Vita-Salute San Raffaele University, Milan) present the results of two experimental studies. The first study aims at establishing whether the offensive component of slurs exhibits nondisplaceability. Data show that the derogatory content survives in conditionals and questions (supporting a pragmatic approach), and diminishes in indirect reports (in line with presuppositional accounts). Surprisingly, the offensiveness of slurs results almost nullified in negated sentences. In the second study, the authors explored the hypothesis that negated slurs were rated as not offensive because the negation was interpreted as metalinguistic.
"Silencing Speech with Pornography" by Laura Caponetto (Vita-Salute San Raffaele University, Milan) closes this section. The aim of this paper is to offer a map of the dynamics through which pornography may silence women's illocutions. Drawing on Searle's speech act theory, the author takes illocutionary forces as sets of conditions for success. According to her thesis, the different types of silencing originate from the hearer's missed recognition of a specific component of the force of the speaker's act. She concludes by suggesting another kind of silencing produced by the failure to acknowledge the speaker's words as serious (*seriousness silencing*).

The final section of this volume, *Emotions and Intersubjectivity – Typical Development and Pathologies*, is opened by the invited contribution by Thomas Fuchs (University of Heidelberg): "Intercorporeality and Interaffectivity". According to phenomenological and enactive approaches, human sociality does not start from isolated individuals, but from intercorporeality and interaffectivity. To elaborate this concept, the paper introduces a concept of embodied affectivity, regarding emotions as a circular interaction of the embodied subject and the situation with its affective affordances. This leads to a concept of embodied interaffectivity as a process of coordinated interaction, bodily resonance, and 'mutual incorporation', which provides the basis for primary empathy. Finally, developmental accounts point out that these empathic capacities are also based on an intercorporeal memory that is acquired in early childhood.
In their invited article "Emotion Dysregulation in Borderline Personality Disorder: A Literature Review", Cesare Maffei (Vita-Salute San Raffaele University, Milan) and Viola Fusi (Vita-Salute San Raffaele University, Milan) propose a review of the literature with the aim of describing the state of the art related to emotion dysregulation in Borderline Personality Disorder (BPD) patients and to illustrate a possible descriptive model. Besides the absence of a general consensus, the authors claim that there is a general agreement in understanding emotional dysregulation both as affective instability and as a frequent recourse to dysfunctional regulation strategies. The latter is what emotion dysregulation is from an operational point of view. The descriptive model explained in this paper is a possible way to bridge the conceptual and operative views of emotion dysregulation in BPD.
The paper by Flavia Felletti (University of Barcelona) – "What Autism Can Tell Us about the Link between Empathy and Moral Reasoning?" – discusses the relationship between empathy and moral reasoning among people with autism. By discussing the specific deficit that can be found in autism and by providing some experimental data, the author supports the thesis that, unlike typically developed, people with autism have difficulties in perspective-taking.

Moreover, she concludes that studies on autism do not help to assess the influence of the affective components of empathy on moral reasoning.
In her paper "Extended Affectivity as the Cognition of Primary Intersubjectivity", Laura Candiotto (University of Edinburgh) explains why the distributed cognition model is the most convenient to understand the collective and the subjective dimension of extended affectivity. She proposes extended affectivity as the cognition of primary intersubjectivity after considering both the primordial affectivity approach and the extended emotions theory. She claims that the novelty of extended affectivity as the cognition of primary intersubjectivity consists in the recognition of the protocognitive valence of affectivity.
The last paper of this section – "On the constructive role of conflicting emotions: The case of early mother-child interaction and its relevance for the study of social behavior" – by Roberta Patalano (Parthenope University) defines a preliminary basis for a dialogue between philosophy and psychoanalysis on the topic of emotional conflict. The author argues that the interaction between mother and child in the latter's first year of life represents a privileged vertex of observation for the positive effects that can be produced by coping with emotional ambivalence, both on the quality of the relationship and on the development of the child. Furthermore, tolerance for emotional conflict not only contributes to the development of the Self in the infant, but it also favors the acquisition of prosocial attitudes, such as the capacity for concern, authenticity, and creativity.

Acknowledgments

We are grateful to all those who made this experience possible. First, we wanted to thank all authors for their presentations, their papers, and the thought-provoking discussions we manage to have during the School.
Second, we thank the members of the Scientific Direction – Claudia Bianchi, Roberta De Monticelli, Francesca De Vecchi, and Roberta Sala – for the great opportunity they gave us helping them organize the School and editing this issue. They trusted us and we are very grateful for that.
Third, we need to thank also all our reviewers for their availability and their amazing job.
Fourth, we are grateful to the administrative personnel who helped us in the organizational process of the School – namely, Clelia Cantoni, Marzia De Masi, Valentina Delfino, and Veronica Milani.
Finally, we thank all the members of the Organizing Committee – Bianca Bellini, Laura Caponetto, Chiara Cozzi, Simone Carrus, Margherita Isella, and Maria Russo – for all the effort both before and during the School and for making it such an amusing experience.

REFERENCES

Bornemark, J. (2013), "The Genesis of Empathy in Human Development: A Phenomenological Reconstruction", *Medicine, Health Care and Philosophy*, Vol. 17, No. 2, pp. 259–268;
Frijda, N. H. (1986), *The Emotions*, Cambridge University Press, Cambridge/New York/ Melbourne/Paris;
——— (1988), "The Laws of Emotion", *American Psychologist*, Vol. 43, No. 5, pp. 349–358;
Gallagher, S. (2001), "The Practice of Mind. Theory, Simulation, or Primary interaction?", in *Journal of Consciousness Studies*, Vol. 8, No. 5-7, pp. 83-107;
Goldie, P. (2000), *The Emotions: A Philosophical Exploration*, Oxford University Press, New York;
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001), "An fMRI Investigation of Emotional Engagement in Moral Judgment", *Science*, Vol. 293, No. 5537, pp. 2105-2108;
Hofer, M. A. (1981), *The Roots of Human Behavior: An Introduction to the Psychobiology of Early Development*, Freeman, New York;

——— (2006), "Psychobiological Roots of Early Attachment", *Current Directions in Psychological Science*, Vol. 15, No. 2, pp. 84–88;
Kohut, H. (1977), *The Restoration of the Self*, International Universities Press, Madison;
Nussbaum, M. C. (2001), *Upheavals of Thought: The Intelligence of Emotions*, Cambridge University Press, Cambridge;
Prinz, J. (2004), *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press, New York;
Schore, A. N. (1994), *Affect Regulation and the Origin of the Self: The Neurobiology of Emotional Development*, Lawrence Erlbaum Associates, Inc., Publishers, Hillsdale;
Stern, D. N. (1985), *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*, Karnac Books, London;
——— (1990), *Diary of a Baby. What Your Child Sees, Feels, And Experiences*, BasicBooks, New York;
Whiting, D. (2011), "The Feeling Theory of Emotion and the Object-Directed Emotions", *European Journal of Philosophy*, Vol. 19, No. 2, pp. 281–303;
Zahavi, D. (2014), *Self and Other: Exploring Subjectivity, Empathy, and Shame*, Oxford University Press, Oxford.

SESSION

SESSION 1

THEORIES OF EMOTIONS

DIETER LOHMAR
Universität zu Köln
dieter.lohmar@uni-koeln.de

EMOTIONS AS A MULTI-MODAL SYSTEM OF REPRESENTATION IN NON-LINGUISTIC THINKING

abstract

Emotions can carry meaning, evaluations, knowledge, experience and sometimes also plans for the future, thus they can function as part of non-linguistic thinking. In my contribution I would like to establish the function of emotions as an important part of an effective system of thinking and deciding that does not use the concepts of language but visual scenes combined with feelings.

keywords

non-linguistic thinking, daydreaming, meaning of emotions, multi-modal use of emotions, the scenic-phantasmatic system of representation, deciding in non-linguistic thinking

Phenomenology and Mind, n. 11 - 2016, pp. 20-30
DOI: 10.13128/Phe_Mi-20103
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

First I should make clear that my view on emotions in human mental life is quite exceptional: I will consider the function of emotions in the framework of non-linguistic thinking. In this view emotions are usually connected with phantasmatic scenes of a state of affairs or a connection of events, and they carry sense, evaluations, experience and sometimes also plan and a look on the future. In a sentence: Emotions are an important part of a system of thinking that does not use the concepts of language, but is somehow older than language.
Now I have to explain two items: (1) Why does it make sense to speak about non-linguistic thinking and (2) why should we take into account the age of language in this concern? Well, first because it is important to know that language is not as old as we tend to think. There is hard evidence from palaeoanthropology that language is not more than approximately 120,000-150,000 years old. The evidence for this is that today all living humans (homo sapiens sapiens) have a characteristically large-sized tongue bone that can be used to form extensively vocalized language expressions. In comparison to this large tongue bone, the corresponding bone in non-linguistic monkeys and also in Neanderthal hominids is quite small. Today, it is widely supposed that the articulation of spoken language is strictly correlated with this anatomic characteristic. On the other hand, we know from studies in the lifestyle of early hominids that there has to be a function of thinking and conceiving of future plans and the past during the hominid's lifetime since the offspring of *homo erectus* (1.8-2 million years ago). But this way to think cannot be linguistic, but nevertheless it has to solve important tasks. So quite a big gap of time emerges in which hominids have to be able to think but are not yet able to use language.
The idea of investigating non-linguistic systems of thinking thus tries to solve this enigma of a function that our hominid ancestors used to have but that is for us living *homo sapiens sapiens* today somehow difficult to detect. The alternatives are obvious: Either the function of non-linguistic thinking has vanished completely with the rise of language, or it is still there but we cannot recognize it. A further alternative is: Such a phylogenetically old system of thinking might well be conserved in our consciousness as a *redundant system* that solves the same problems as our language system of thinking does, but it still remains functioning and it immediately steps in for the function of thinking in situations where we cannot make use of our standard language-based thinking.

1. The principle idea of non-linguistic thinking

I will now list some topics that a non-linguistic thinker must be able to think about - simply imagine you were a chimpanzee in a group of others: 1. *states of affairs* of special parts of the world, for example whether a banana is dark yellow, ripe and tasty or green and unripe, 2. the *character*, the *motives*, the *social status* and the *plans* of other persons of my group, for example if there is another member of the group with a higher status in hierarchy and more privileges, he will easily take the banana I found first. 3. the *meaningfulness* of such states of affairs, for example what relevance the state of ripeness has for my own life, 4. this implies also an *evaluation* of an event or a state of affairs and this evaluation could be limited to (solipsistic) values I can judge on my own and on the basis of my experience alone, say, ripe bananas are tasty and I like them, 5. Perhaps also 'objective' values that are shared by my community and thus have a more objective character, 6. the *certainty* of my knowledge in single cases, i.e. whether I am very certain about a fact so that it will lead my actions in the future or not, 7. the location in *time* when my conception of state of affairs has been true or will be true (i.e. whether it is my wish or a memory I am thinking about): yesterday, just now or in the future. - This list may be not exhaustive.

I am quite sure that I will not be able to discuss all aspects of this list of necessary themes of a thinker who is not able to use language. What is quite obvious is that we humans are able to conceive of all these topics using language. But I will try to convince you that we humans are still able to think about most of these themes without the use of language. For this reason, we are also allowed to suppose that intelligent animals might use the same means of thinking. And in my view, feelings do play quite an important role in this non-linguistic thinking.

Alternatives for symbolic representations in non-linguistic thinking

What are the non-linguistic means we might use for symbolic representation of state of affairs? I will name three types (This list does also not claim completeness): 1. *Codified national gesture languages* (German gesture language, ASL, etc.). 2. Non-codified gestures together with mimics and pantomimes (this is a *hand & foot communication system* I will come back to later.). These first two modes are usable for communication as well as for thinking. 3. A *system of scenic phantasma* consists of series of pictures in our imagination sometimes uniting to a kind of video clip of past and future events combined with feelings. It is suitable for the representation in solitary thinking but it cannot be used for public communication. Sometimes we tend to identify scenic phantasma with daydreams.

Generally, the connection between language and thinking is not as tight as we tend to believe. Not only can we express our insights in different languages, but we can also think in a language other than our mother tongue. Most of us are familiar with the following experience: After spending some days in a foreign country where a foreign language is spoken with which we are familiar, *our thinking takes on the form of this other language*. This example shows that the level of language is only on the surface of the whole phenomenon of thinking.

But do we need a symbolic representation at all to think about our cognition and our options to act in the future? I think so because seen from the phenomenological theory of cognition, we can hold on to the intuition of states of affairs (categorial intuition) only for a short time. After this we must have a symbolic medium to hold on to the contents of our cognition, but it must not be language.

The access to a symbolic carrier of a conviction is the presupposition for the three essential performances of thinking: (1) the ability *to awaken* and *to retain* in mind the same object of cognition; (2) the ability to *engender* other cognitions from this one; (3) the ability to *manipulate* our future possibilities. These central performances allow me to think about the possible future of an object or event in different situations, ponder possible consequences, obstacles and alternative solutions of problems. Essentially, thinking is an active treatment of the contents of our cognition.

Thus thinking *must have a medium of symbolic representation,* but this medium need not be language. Yet language gives us a hint to the most important feature of such a system of symbolic representation: I must be able to produce the material carriers of symbols at any time; for example, I must be able to produce spoken or written words at any time either in public speech, or in inner speech. I am only able to think if the symbolic carrier is ready at hand all the time. This carrier must achieve its meaning in a meaning-giving act based on the intuitive cognition. This is true for language and for all non-linguistic systems of representation. In this regard, the use of non-linguistic symbols follows the pattern of Husserl's *theory of meaning*[1].

Thus we may conclude what we already know: language *is* a usable carrier of cognitive meaning; it makes thinking and public communication possible because I can speak loudly any time. And in regard to inner thinking, I can let my inner voice function as the carrier of thinking. But our conclusions can go also beyond this trivial insight because I now know at least the minimal requirements of symbolic systems useful for thinking: I must be able to produce the carrier of symbols at any time - either in inner or outer sensibility. Thus internal carriers of meaning that allow for thinking but *not* for public communication can also occur. And there may also be carriers of symbols that allow for both, such as language, gestures and pantomimes. In the next part, I will concentrate on the scenic-phantasmatic system of representing cognitive contents that belongs to the first class. In my view, this is the basic mode of our non-linguistic thinking.

2. The scenic-phantasmatic system of representation

The first problem we have to solve is to understand how we are able to think of states of affairs without using language. In this concern, I have to shortly characterize the scenic-phantasmatic system of representing a state of affairs or the follow-up of events. The scenic-phantasmatic system of representation is quite simple because I only have to imagine vividly a state of affairs (it is not yet determined whether it is a present, past or future state): I "see" a ripe banana with the help of phantasma.

This phantasmatic picture makes the banana appear to me in full color and perhaps also with the slight smell of banana. I like ripe bananas and a felt desire to eat the banana comes up in me, raising more positive *feelings*. - Then perhaps I imagine that another person comes into the scene, grabs the banana and eats it before my eyes and upon seeing this, I feel great *disappointment* perhaps mixed with *fury*. - And there are even more feelings connected with this scene because this other person might be a highly recognized member of my group with high privileges, and a feeling of *respect* arises in me. - Then I also feel *anxious* about my sudden wish to ignore his privileges and to start a fight with him: I may try to grab the banana from him. Imagining my possible act of disobedience (opposing the rules that guide our community) a feeling of *fear* accompanies my imagination. Imagining the following sanctions of some other members of the group let me fear future pains. - So we see that there is grief, fury, and anxiety connected with the same basic situation and its possible developments. Each of these feelings is representing different aspects and is related to *different future options and resulting events* coming into my mind spontaneously. And we also understand that beside the evaluation in the accompanying feelings, there is already an indication of the location in time and the follow up of the imagined events. Fear can only be connected with a future event.

1 Cf. E. Husserl: *Logische Untersuchungen*. Husserliana XIX, Den Haag 1984. I would like to thank Andrew Krema for his kind help with the English text.

By this short story of my vivid imaginative ideas about *real* or *possible* states of affairs[2] in a non-linguistic mode, I wanted to show you that humans are using this method of representing future and past events and states of affairs by using pictorial presentations and scenes that run in my fantasy - and all of this without language. And you will also have realized that feelings are accompanying these fantasies adding important elements of sense concerning the respective evaluation, relevance, location in time of this real or imagined events. - We might even think that we have already detected a full-blown system of representation for the most common but quite basic events of our everyday life.

It seems to me that in series of imaginative pictures and in *daydreaming,* we are using scenic phantasma to express our knowledge of states of affairs as well as to think about our wishes and fears - generally they function as representations of cognitive contents connected with practical and evaluative components. It is always a state of affairs that we wish for or are in fear of. But we do not only express our preferences, our urgent wishes and our views of the state of affairs by these means. It turns out that the scenic-phantasmatic system can also be a kind of response to a (real or possible) problem. Sometimes we may find even a *mental action*, a mental manipulation of the problematic situation might lead to a solution of something which until now was unthought of. My thesis is this: *Series of phantasma and daydreaming are an old mode of thinking still working in our consciousness*[3].

In my view in daydreams, we are playing out possible solutions to a problem, i.e. we are mentally testing our options, their usefulness for a solution and their respective consequences. This "life" of scenic phantasma constitutes a great and important part of our conscious life. Some examples: Worries about urgent challenges and the effects of events or uncertainties concerning possible developments that make us sleepless at night. There are many fantasies of having success. In these scenic episodes of our conscious life, the linguistic expressions emerge in the background in favor of pictorial elements. Thus we realize the meeting of past, present and possible future events in our scenic phantasma, but we still have to find out how this thinking proceeds[4].

2 It may seem that the modality of the events is still not determined, but it may be that in this regard feelings do have a say too, for example the secure feeling about a state of affairs may express its reality etc. And it may also turn out that we have to be more precise in describing the connected feelings, for example in the case of disappointment about a possible development we might in fact *feel different* than in the case of a real disappointment.

3 To work out this hypothesis I will have to ignore for a while some other theories about the status of fantasy in daydreaming. There might be objections to our very "reasonable" interpretation of daydreaming from different points of view: From a *liberal-fantasy point of view* our fantasy is usually completely free in the formation of daydreaming and therefore it cannot be of any use when it comes to the "real" problems of everyday life. But some sober reflection and self-observation will convince us that we are not completely free in the formation of our daydreams. - From a *part-part point of view* we might suspect that we are free in the formation of our positive and pleasant daydreams but passive in the formation of our daydreams about lasting fears. This is not the case either: in both cases I experience myself to be bound. - From a *psychoanalytic point of view* we might suppose that all the contents of our daydreams are closely bound to our individual experiences just like our dreams are bound to them. But we will have to ignore these theories about our daydreaming for a while.

4 We might also take a side-view on animals. We know that most highly developed mammals can dream (for example dogs). While dreaming they show first signs of an attempt to act and emotions. We might interpret these phases of their sleep as dream episodes prolonging wakeful states of action and representing aims.

We might therefore claim that a system of representations on the basis of scenic phantasma combined with feelings, is also operative in higher cerebralized mammals up to primates in dreams and wakeful state in the same way as in humans. This claim, however, only indicates an important consequence that stems from my investigations into the systems of representation in humans. But nevertheless this hypothesis about animal thinking is not mere fancy or an arbitrary fantasy because, as the phenomenological analysis reveals, it characterizes an important dimension of our own thinking. Thus through these analyses we might find out in which way we are still thinking like animals. In the present analysis I will not concentrate here on the theme of animal thinking. Cf. for this D. Lohmar (2016), *Denken ohne Sprache*, Heidelberg.

In the simple examples we have also realized that feelings are an important element of the non-linguistic system of representation, functioning in the framework of scenic phantasma[5]. Emotions can easily grant the most important request for a system of representation for we can have them in an actual situation and we can also "produce" them (although not arbitrarily) in the absence of the intuitive situation, i.e. only through imagination. For example, the feeling of fury might move me violently in a certain situation, and the same feeling can also reappear in mere thinking of the same situation later on. In both cases the feeling "tells" me something about the value of this event, it is a part of my inner "expression" that has a certain meaning. In thinking about a nice experience, the pleasant feeling "means" the desirable quality of the event[6].

But when analyzing feelings in non-linguistic thinking, it turns out that the most usual way in which feelings arise is in the context of an intention of an object of thinking. Thus, it has to be an object that is present to me in the scenic phantasmatic way of vivid phantasma before I can combine it with emotions that mirror for example my evaluation or the relevance of the event. Emotions cannot stand alone in this concern[7].

We might realize another difficulty while regarding feelings as part of a symbolic system of representation. It is obvious that in using language symbols for thinking about cognitive contents, we have a certain freedom of choosing alternative wording to express the same cognitive content. We may speak about a nice or pleasant outcome, or we might speak about a necessary or an unavoidable outcome referring to the same state of affair. When using feelings in the scenic-phantasmatic system in thinking about an outcome in our actions and its high value, we simply have no choice: a pleasant feeling accompanies the idea of this resulting state. In non-linguistic systems, we cannot use an alternative "expression" to characterize my valuation[8].

We see that we should not expect the same characteristics in a system of non-linguistic representation as in the case of language. The two systems of representation are phylogenetically at a great distance from one another and they are using quite different semantics[9].

5 In my view we can not interpret emotions as an independent system of representation because we always have to presuppose another kind of representation in which we have in mind things or (real or possible) events that are object of feelings.

6 We might suppose that also most animals have feelings as part of their system of representation. It makes no sense to have objects and properties of them if you don't have feelings to evaluate this objects and events, this is the way to actively make use of our experiences with them.

7 This is to a certain extend an opposition to Heidegger's claim that moods are a way in which we relate to the world as a whole. But also in this case the world is also an object of my intention.

8 Beside this in non-linguistic thinking we are not obliged to use for example a more neutral expression of our feelings like we do in public language to meet the standards of our society. Think about the case where the pleasant outcome is in fact an accident that hits a personal enemy. We feel pleasure about his accident but it might not be accepted by the community to speak about it in terms of pleasure. Therefore I choose another wording somehow hiding my true valuation. But obeying this kind of rules is specific to systems of representation useful also for communication. Only here the rules of "good manners" are working. In my private, solipsistic non-linguistic thinking I do not feel to obey these rules. Thus in a certain concern non-linguistic systems are more truthful to what we really think - but in other concerns they do not have to be so truthful. Cf. for this D. Lohmar (2012), "Psychoanalysis and the logic of thinking without language. How can we conceive of neurotic shifting, denying, inversion etc. as rational actions of the mind?", In Lohmar, D., and Brudzinska, J. (eds.), *Founding Psychoanalysis. Phenomenological Theory of Subjectivity and the Psychoanalytical Experience*, Springer, Heidelberg, pp. 149-167.

9 Cf. for this aspect D. Lohmar (2016), *Denken ohne Sprache*, Heidelberg, Kap. III, 3.

3. The multi-modality of feelings in the non-linguistic system of thinking

There are more important facts about the function of feelings in the non-linguistic system of representation and thinking in humans (and perhaps also animals). I have already announced these aspects in the heading of my paper by the word "multi-modality" of the function of feelings. What we have already realized is that feelings can help in many places of this life of imagination to fill in elements of sense that we cannot represent simply in scenic-phantasmatic pictures - like evaluations and meta-cognitive elements of sense like security or insecurity of our knowledge. We have to be able to represent these elements of sense somehow in our non-linguistic thinking; even if they are a more sublime, they cannot be sensed by one of our senses. I will discuss some of these aspects in the following.

The aspect of relevance or meaningfulness for my life is represented in the feelings of liking and disliking, grief, and promise. For example, when I see the banana, I feel joyful about the possibility of eating it, I am "happy in advance" concerning this possibility. Thus we have to be attentive that being happy in advance (in german: *Vor-Freude*) feels different than simply being happy now. And the same is true for "feeling happy afterwards" (*Nach-Freude*).

Let us think briefly about *security* or *confidence* concerning a possible solution to our problems. Both aspects are on the level of metacognition for they do not concern the content of cognition but rather a meta-quality like the knowledge about the source or the reliability of this knowledge.

Think about the problem that you have a serious lack of money. Different solutions then pop up in our mind: For example, winning the lottery will easily solve the pressing financial concerns, but it is unlikely to happen and does not give me a feeling of *confidence*; working hard or suffering for some time from certain hardships will work as well, and this idea gives me much *more confidence* in its success. - This shows clearly one function of feelings in non-linguistic modes of thinking: Feeling is realistically adjusting to the chances or probability of the effect of our actions. Therefore, daydreaming should not be interpreted as an evasive regression to a childish mode of handling problems only in fantasy[10]. There is a strong realistic trait in daydreams and - surprising enough - it is hidden in feelings[11].

This opens up an easy way to understand the meta-cognitive abilities of many animals. Up to now there are only a few insights in the meta-cognitive abilities of primates, but this kind of empirical research has only been done over the last 10 years. With a view to the role of feelings in non-linguistic thinking, we immediately understand why animals can also think about such meta-properties of our intentions in states of affairs. The only thing we or other animals need is a representation of the state of affair itself - this can happen in scenic phantasma - and additionally a feeling of security that accompanies this idea of a state of

10 We might also stress that non-linguistic modes of thinking in the scenic-phantasmatic system are not as quick and effective as linguistic modes of thinking; it will always take some repetitions to find a way to solve the problem, even if the scenic representation of an event is running in a compressed speed-mode.

11 In the context of phenomenological analysis it is sometimes discussed whether evidence is a kind of feeling or not. What is quite obvious is that in the constituting higher level insights we have to have a means also to "conserve" somehow whether an intention has been evident to me and in what degree there was evidence. For example: If we take into concern the building up of knowledge in science and in our life-world then we usually have to relate in a way to simple knowledge and its evidence in one ray. In the *Logical Investigations* Husserl speaks of nominalization to denote the condensed intention to a cognitive content after having intuition of the state of affairs in categorial intuition: We first realize that the breaks of the car are defect and conclude: "This is dangerous" referring to this state of affairs by using "This". And in using the word "This" for the state of defectness of the breaks we also have to have a kind of surrogate for the evidence we have had. Therefore you see that feeling is not only a necessary element in non-linguistic thinking but it also accompanies our language use (but without realizing this). Therefore we may conclude about the relation of evidence and feeling (1) that evidence is no feeling but (2) certain feelings are used to represent evidences we have had before in the context of the next step of cognition, for example in a conclusion: This is surely dangerous.

affairs. The meaning of this feeling is the conviction that the state of affairs we are thinking of is a fact, that it is real.
There is another aspect of meta-cognition closely connected to the probability of events to happen: It concerns our confidence to act effectively in a certain way. I "think" myself to be capable of some action and this conviction - which is in fact a meta-conviction - is mirrored in my confidence, in my safe feeling concerning this activity. On the contrary, when I do not feel confident in my abilities to act - for example to climb a wall - I feel helpless and depressed when thinking about my planned action.
Feelings are not a simple element of conveying sense in non-linguistic thinking; sometimes the contents they have may be quite complex. We see this in the case of social feelings like pride or shame. In being ashamed, we are in shame *of something* - this is usually a past action or something related to my body - I may have curled hair, dark skin, a clumsy figure or my clothes do not fit the lifestyle standards etc. - or my *personal history* - my father may have been a drunk, or in prison etc. And, I am ashamed *in the eyes of a community*, a group of people that share a valuation of my person. Therefore, social feelings usually connect a complex net of relations of the valuation of history and personal relations etc. This mirrors generally one of the big advantages of feelings: they can represent very complex relations.
We cannot see time, nor taste it, nor sense it in any other way. Therefore time is difficult to indicate in the scenic-phantasmatic system besides being indicated in the form of extended episodes that are structured in time - a kind of narrative story. We might think of time passed in seeing flowers blossom in spring, or snow on the hills may indicate winter etc. but it is difficult to indicate whether events I am conceiving in my scenic phantasma are past or future events without the help of additional information. And in this situation, emotions also do their work in the framework of non-linguistic thinking by indicating past or future.
To realize this, think of a nice event that is going to happen this afternoon (i.e. in the future) or yesterday (i.e. in the past). You will notice a difference of feeling in these two cases - or perhaps you are not so sure whether a difference is felt? If we were to take language concepts as a clue in this concern, we might think of the differences of naming: pre-joy and past-joy (in German: *Vor-Freude* and *Nach-Freude*). Try to find out what is indicated by these words. And try to be attentive to the differences in your own experience: Is it more arousing to think of the future event - is to think of the past event exciting in the same way? Definitely not. But you should also be attentive of the contents of your thought: In thinking about a joyful event in the past, we may easily be tempted to skip over to a possible repetition of the event in the future and the same exciting arousal of joyful expectation might accompany our thought. - Thus we realize that emotions can also indicate the place in time and time relations of events[12].
It seems difficult to imagine a scenic image of the character of a person and of his or her probable behavior towards me, especially within complex constellations with others involved in action. But scenic phantasma offer a simple solution to this apparent difficulty. In remembering a brutal former classmate, I see his face looking at me with evil eyes, with clenched fists, and ready to give me a beating. But this "image" is not simply an image of him; it is a very characteristic scene within which I am present, writhing with pain from his beating and in fear of his further beatings, and in the background, there is a group of "friends" not helping me. This scene presents central aspects both of his character and of his future behavior within a social context. -- And there is also the *felt aversion of him towards me* that is

12 It goes without saying that this is an important part of the function of syntax. But this should not mean that non-linguistic systems of representation needs a syntax, it is only to hint that all systems of representation have to perform certain general functions, but they realize them in different ways.

present in his facial expression. I *feel* that he hates me. The co-performing of his hate is not to identify with self-hate, but nevertheless, feeling is the only way of representing his deep motives.

But the scenic presentation of the attitude and behavior of a person need not be so one-dimensional as in this case, since there are normally multiple facets of the character of other persons that we are able to present. Thus the question arises: How can I think a multitude of (changing) attitudes in a scenic mode? Now, think of a colleague with whom you work together successfully in most cases, but who occasionally appears to have an air of high-nose arrogance. Both "faces," i.e. both aspects of his character, may be represented in a scenic phantasma, one after the other, or, even, as *mixed in a changing way*, which results in an uncertain base for your plan making. The modal character of *possibility* and *uncertainty* is thus also presented in the changing and merging faces of your colleague. We might even interpret this changing image as a non-linguistic form of the logical "or." Additionally, his attitude towards other persons and his preferences to act in a changing situation may be represented in a short but eloquent side-view of others.

The value and the usefulness of objects are also reflected in feelings. And as we know that the reliable qualities of objects can change, this may be also reflected in feelings. For instance, if I own a car that usually breaks down and thus has to be towed off and repaired, the characteristic scene within which I am *positively excited* about my car is modified and converted to one that is negative. The emotional aspects of this bad experience are mirrored in feelings: I no longer imagine the car with the joyful expectation of reliable use, but with the *cheerless expectation* of future harm, expense, and inconvenience.

In the non-linguistic system of thinking in the mode of scenic phantasma, feelings have another important role that is quite difficult to understand because it deals with the big problem of non-linguistic thinking about general ideas. General ideas are what we speak about in the language system as concepts, ideas of not only one object but a group or class of objects, like horse, man, animal, living creature, fairness, justice etc. In the non-linguistic system of representation, there is a rather easy method of thinking of *low-level general ideas* like horse or man in using a vague visual phantasma of what we think of. Vagueness is one useful method to think about low-level general ideas without using language. But this method does not lead up to a high point on the mountain of generalities; already with "living creature," we exceed the limits of generalities that we can think of with this means of vagueness.

There must be other means to "think" about *high-level general ideas* without using language concepts. One of these means is using objects of our experience in an exemplary way. We might therefore call this method *exemplary semantics*, and we find it, for example, where we are trying to think about unlimited generosity, benevolence or moral integrity in a person, then suddenly a picture of our grandfather comes into our mind. In this situation, he represents unrestricted benevolence - but in exemplary semantics - and this is a general idea of higher order, i.e. an idea we cannot represent in the pictorial mode of scenic phantasma because there is simply no "visual side" of the high-level general ideas of benevolence, justice etc.

But also in low-level general ideas, there is a fine stratification of generality that is not easy to represent with the means of greater or less visual vagueness. This difficulty can be easily exemplified by my remembering an embarrassing situation that also has some witnesses, i.e. there were some people around when the event happened. If we try to think about this event, there is the event itself in the center of my scenic phantasma, but there are also these people who come to see the embarrassing event too. Who are they? These witnesses in the background are presented only vaguely if I remember the situation but my feeling tells me additional things about them for I imagine them with a graded *feeling of acquaintance*. I understand that perhaps they were colleagues, neighbours or even friends of mine. But without fully individualizing

them in my scenic phantasma, I only feel that they are not complete unknown to me – perhaps they were neighbours. If the feeling of acquaintance is weaker, it may have been only loose acquaintances, people I have met only once in a while. - In this way, emotions also modulate the generality of low-level general ideas in non-linguistic modes.
We see that emotions can carry a multitude of meanings and functions in non-linguistic thinking and that it will be difficult to reach a complete list.

4. On deciding in "emotional calculation"

Now we may ask: Is the fact that there are so many mixed and concurring emotions found in our non-linguistic thinking about former events and future plans a problem for this kind of thinking, or is it an advantage? It is not so easy to find an answer to this question because there are in fact possible difficulties concerning the extensive use of emotions in non-linguistic thinking. I will discuss them in a minute.
Beside disadvantages, there are also big advantages in the mixture of emotions in non-linguistic thinking. I will first give you a rough idea about the general advantage of deciding following the result of an "emotional calculation", then I will give you a concrete example, and later I will discuss problems arising out of this fact.
"Emotional calculation" has to be used in non-linguistic thinking in situations where there are many different factors that influence our decision. The decisions in everyday life are usually of this kind. Think for example of the decision to choose a certain restaurant. Here we find the factor of quality of food, but this is not the only factor because the ratio between quality and expenses usually lead us to a kind of compromise. Beside this, there may be other factors like the experience that my favourite restaurant is usually overcrowded, so there is a substantial probability that we will not get a table. Besides that, if I am short on time, there might be the factor of time expenditure in reaching my favourite restaurant. - But how are we able to find a solution to this difficult decision? There are factors that influence our decision-making, but these factors cannot really be set against each other in a conceptual way: In what rational relation can we think of quality and price, of probability and time etc.? In fact, we solve problems like this all the time, but the solution is not based on concepts and rational calculations. We come to a solution simply by listening to our emotional answer in posing this kind of complex problems.
Speaking of "emotional calculation" should not imply that this method of relating partly paradoxical motives is somehow a calculation with numbers on a methodical basis. "Calculation" in this context is used only as a metaphor for our ability to reach a decision in the stormy centre of mixed emotional motives.
This kind of mixture of motives is to be found in so many situations of our life that it is easy to see that we only gain our capacity to act by using the "emotional calculus". And there is another advantage of this method of decision-making: It can be done in a second. - I will not claim that this decision is always the best decision we are able to find. It will often point out that there might have been much better solutions. But in the point of view of evolution theory, the mere fact that we are able to decide in a situation with very mixed motives must be understood as an incredibly big advantage.
Besides that, there are advantages of the strong mixture of emotions in the scenic-phantasmatic system of representation, and there are also some disadvantages of founding our decisions on the "offset of different emotions" in an "emotional calculation." One of the very obvious problems is that we have to take elements stemming from very different sources into our "calculation" of the mixture of emotions. I'll give you an example: On the one side, there may be our strong aversion of certain persons, say Peter, who usually starts his day by playing a mean joke on me. Think of the situation when you enter the office after having realized that someone has let the air out of your bike's tires, so you are already suspecting that Peter might

be the one. Then he welcomes you with a broad grin in his face, and you immediately know: It was him, i.e. you are sure without any further examination that it was him, even though you heard children laughing in the bushes in the background near your bike. In cases like these, we might speak about "prejudices", and this may be true if we are only language-using thinkers, but in fact there are strong emotions stemming from completely different sources, and we do not hesitate to rest our judgement and the further actions on this emotional information. Thus you see immediately that emotions stemming from our wishes concerning future events, our fears and aversions and on the other side emotions representing the security of former insights etc. may be calculated together as if they were the same "currency" and deserve the same respect.

In the end we see that we can consider emotions quite different from the standard view, and there are advantages to look from a completely different angle at emotions as a central multi-modal element of a non-linguistic system of thinking. But as we are now informed about the great variety of information and valuations entailed in the condensed emotional attitudes we experience in our everyday life, we may easily run into radical questions about the relation of the two systems (Language and non-linguistic thinking). I see in principle following alternatives: 1. The non-linguistic system is only a non-functional redundant system of our consciousness, and real thinking takes place only in the mode of language concepts (primacy of language); 2. Both systems work parallel in our thinking and they do not influence each other (parallel systems of equal performance); 3. Our language-based thinking is only a kind of second, symbolic form of the fundamental non-linguistic system of thinking. I should confess that this is the thesis I am sympathizing with most (primacy of non-linguistic system of thinking).

This third alternative implies that our rich life of highly condensed scenic phantasma, accompanied with multi-modal emotions, is the basic way of thinking in everyday concerns. Language is only reformulating the problems that are presented in very condensed forms in this non-linguistic thinking, but it does not really have a productive impact and makes only a minor contribution to our thinking. Language only "translates" what we have been thinking and deciding before in non-linguistic ways, and thus is only a rather superficial part of the whole process of thinking on everyday topics.

Nevertheless language has the merit to be a means of communication so that we might be able to speak about and discuss items of some importance with others. By this, we reach a new level of constitution in intersubjective thinking about states of affairs in the objective world. It seems to us that only by communication we can reach this level of common opinions on common topics. And it helps us to make up our mind about topics like fairness, justice and other important topics that can only be conceived in high-level general concepts. But it seems to be difficult to find a decision on this level of generality without going back to the everyday intuitions that guide our actions.

In the end, we had quite an interesting journey into the deep subjective core of a non-linguistic creature. And in fact, we realized that we humans are quite like this, a big part of our conscious life proceeds in a non-linguistic mode, and a big part of our everyday decisions rests on the rather strange form of calculation in emotional currencies. I do not tend to interpret this way of deciding as "irrational" (only following the prejudice that emotions are irrational); it is a path to getting back to finding the special rationality of creatures that are also able to think in non-linguistic modes.

ROBERTA DE MONTICELLI
Vita-Salute San Raffaele University
demonticelli.roberta@unisr.it

IS THE PERCEPTUAL MODEL OF EMOTIONS STILL A GOOD COMPETITOR? A SMALL PHENOMENOLOGY OF FEELING

abstract

This paper defends a version of the Perceptual Model of feeling, according to which feeling is the mode of presence of axiological aspects of reality, or values. After considering the objections to this Model (§1), it presents an intentional analysis of feeling as the core of all emotions, more generally of all the phenomena of affective life, for which it proposes a taxonomy and provides in addition a dedicated phenomenology of feeling consciousness and of the degrees of intuitive cognition of values which may be distinguished in it (§2 and §3). Finally, it reconsiders the objections to the Perceptual Model in order to make sense of them, showing the limits in which they are valid and how to overcome them (§4).

keywords

feeling, emotions, perceptions, values, phenomenology

Phenomenology and Mind, n. 11 - 2016, pp. 32-47
DOI: 10.13128/Phe_Mi-20104
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

The Direct Access Thesis, sometimes referred to as "the Perceptual Model of emotions", is the claim that

(P) Feeling is essentially a perception of the value-qualities, whether positive or negative, of things.

This claim, which was originally put forward by classical phenomenologists (Husserl 1900-1901, Scheler 1916, Stein 1917, Geiger 1921, Hildebrand 1921-1922, Hartmann 1926 – to name but a few of them), has garnered attention more recently among analytical philosophers as well (Mc Dowell 1985; D'Arms and Jacobson 2000; Mulligan 2010; Tappolet 2000, 2011; Goldie 2004). Several critics have also raised objections against these more recent incarnations of the idea (Dokic and Leymarie 2013, Deonna and Teroni 2012, Bagnoli 2011-2013). Their objections can be partitioned into at least two classes. Some objections are concerned with the differences between emotional and perceptual mental states and contents, while other objections are concerned with the issue of value and normativity. According to the first sort, emotions or emotional episodes *are not like* perceptions in all the respects in which a mental state can be analyzed. They differ, in particular, in their phenomenology and their intentionality. According to the second sort of objection, emotions cannot play the epistemic role that the perceptual model would assign to them, namely, they cannot be a source of evidence for value judgements in the way that ordinary perception can be a source of evidence for judgements of fact.
This paper addresses the first sort of objection, arguing that recent criticisms of this variety are perfectly justified as long as (P) is presented as a straightforward (analogical) explanation of emotions[1]. Emotions are *not* like perceptions. What *is* in some sense like perception is the core of emotions as well as of any affective state, namely *feeling.* As in its classic phenomenological rendering, (P) will be shown to apply *to the feeling core* of any emotional experience, "bracketing" whatever aspect of it exceeds pure feeling. Though showing this

1 The focus of my presentation at 2016 Spring School was the other sort of objection, those addressing phenomenological cognitivism concerning values and rejecting the value epistemology embedded in the perceptual model. Yet, since a corresponding paper is already in print for another journal (De Monticelli 2015), I prefer to shift the focus to the first sort of objection for this paper.

would require a full-fledged phenomenology of feeling, this paper will only lay out some basic points about such a phenomenology.

By a "full-fledged phenomenology," I mean something more than what is currently understood (among Anglophone philosophers) by "phenomenology", which is commonly taken to refer to the analysis of "phenomenal consciousness," leaving intentionality for separate treatment. A full-fledged phenomenology of feeling is supposed to describe the peculiar intentionality of emotional life in its essential features, by illustrating the specific intentional operation of *feeling* as lived experience of reality.

To get this important preliminary point, a terminological remark may be helpful. The word "experience" can be used in two senses; a broad one, designating *what-it's-likeness,* more or less equivalent to the German *Erlebnis,* and a strict one, designating a mode of *immediate presence of reality* as an infinite source of information, more or less equivalent to the German *Erfahrung.* Only on this second rendering does it make sense to say that we "learn" from experience, i.e., that we can always *further explore* the given and *correct* illusions and mistakes. That is, in experience, thus understood, we may get reality *right* or *wrong.* No such immediate hold on reality figures in remembering or thinking, let alone day-dreaming, where veridicality is not even an issue. On the other hand, day-dreaming or remembering, and even propositional thinking, do have a what-is-likeness of their own, after all, as a number of authors have recently maintained (Baine and Montague 2011). Yet, emphasizing the qualitative character of these states has nothing to do with underlining the crucial distinction between *Erlebnis* and *Erfahrung*, because no lived experience is *an encounter* with reality.

Feeling, like perception and unlike day-dreaming, is a way of encountering reality, or so I will argue. This paper offers a defence of (P), on the basis of a full-fledged phenomenology of feeling. A full-fledged phenomenology of feeling is an analysis of experience in the strict sense, providing a description of what is distinctive about *feeling* consciousness as a type of *direct* presence of objects to a subject. Such a mode of presence is what Husserl used to call an "*originally giving" consciousness,* a type of experience best exemplified by (external) perception. Yet external perception is by no means the only instance of *Erfahrung* or originally giving consciousness. Further instances include, indeed, feeling and that direct mode of social cognition that phenomenologists usually call empathy.

I shall first very quickly summarize some objections of the first class to the "perceptual model of emotions" (§1), then propose a brief phenomenology of feeling enabling us to make sense of those objections (§2 and §3) by letting us see their source and the limits of their validity (§4).

1. Objections to the Perceptual Model

According to the first class of objections, emotions or emotional episodes *are not like* perceptions of the value quality of things. Emotions, the objection runs, are not to value qualities as visual perceptions are to colours or auditory perceptions are to sounds. As put forward by – for example – de Sousa (1987) or Tappolet (2000), the perceptual theory of emotions would in fact assert this very analogy:

> Being afraid is to perceive danger, being sad is to perceive a loss, much as having a visual perceptual experience is to perceive, say, a red circle. To the adherent of this theory, we do not need to deploy the concepts of danger or loss in order to experience, respectively, fear and sadness, any more than we need to deploy the concept of a red circle to see one (Deonna and Teroni 2012: 67).

There is a weak sense in which this theory is more than an analogy. It is supported by an analysis of the *intentional contents* of perceptions and emotions and of their common

properties. So, perceptions and emotions have the same mind-to-world direction of fit. In this respect, they are like other cognitive states such as beliefs, and unlike the typical world-to-mind direction of fit of desires and intentions, or conative states. This fact clarifies the difference in satisfaction conditions enjoyed respectively by cognitive and conative states, or cognition and will. There is a further experiential feature distinguishing, within cognitive states, perceptions and emotions from beliefs and judgements and capturing what we may call the character of *impression* (Hume) or "being struck" (Mulligan 2009) characteristic of *direct* or *immediate* cognition. Perceptions and emotions share causal self-reference, i.e., they are about what is (experienced as) causing them (redness, danger).

On the side of *phenomenal consciousness*, the analogy can be spelled out in terms of the lived "affection", the state of being "affected" or "touched", of suffering as it were the impact of reality, through the bodily feeling – for example in sweating or trembling - in which the emotion is lived, much as seeing can be felt in dazzlement or hearing in near-deafening loudness.

On both sides, of course, elaborating the analogies immediately brings out the differences. This will lead us directly to the critical objections against the perceptual model.

First of all, one may wonder *which* model of perception we are referring to. Alva Noë has convincingly contrasted two basic attitudes toward perception, giving rise to two opposite ways of "modelling" it: the "scientist's" and the "artist's" attitudes.

> Some years ago I was talking with an artist. He asked me about the science of visual perception. I explained that vision scientists seek to understand how it is we see so much – the colourful and detailed world of objects spread out around us in space – when what we are given are tiny distorted upside-down images in the eyes. How do we see so much on the basis of so little?
> I was startled by the artist's reply. Nonsense! He scoffed. That's not the question we should ask. The important question is this: Why are we so blind, why do we see so little, when there is so much around us to see? (Noë 2015: xi).

By and large, these attitudes correspond to the naturalistic-scientific (not the "natural" or pre-theoretic) one and to the phenomenological one in the language of classical phenomenology (Husserl worked out the similarity of the aesthetic and the phenomenological attitude at length: Husserl 1998: VII, 133 ff.). Not surprisingly, the "artist's" perspective takes for granted that perception is or can be experience in the sense of *Erfahrung,* not just of *Erlebnis*. The exercise of our senses would be full of adventure and discoveries, surprise and amazement, if we weren't most of the time trapped in our perceptual routines, organized by our practical needs and habits, making us "blind" to the richness and intricacy, the charm and horror, the novelty and volatility of any given perceptual scene (indeed, the artist is supposed to rouse us from the dogmatic sleep of *aisthesis* or sensibility).

The criticism of the Perceptual Model of feeling depends partly on adopting a model of perception based on the naturalistic attitude, for example, the attitude common among psychologists and neuroscientists. From the standpoint of psychology, objects are only envisaged as stimuli, i.e., what triggers experiences in the broad sense of the term – qualia of any sort, but also "mental representations", "images" – all that transpires in the theatre of consciousness. In this language, seeing things is exactly like seeing pictures of them: real pictures actually are externalized retinal images. This is what Noë rightly calls the trigger-experience conception: "As if things just cause the lighting-up of experiences inside our heads" (Noe 2015: 97). "Experience" cannot but be meant in the broad, weak sense of *Erlebnis* here, and not in the more appropriate and strong sense of an encounter that a whole

embodied and active person has with whatever presents itself in its surroundings, in the flesh, demanding attention, exploration, investigation, and interaction.
Can one possibly perceive the golden glare of the Pisan Hills at midsummer time, without perceiving the soft lines of a Tuscan horizon, and the orderly sweetness of such a landscape? "Perceiving" means here an exercise of all the senses or sensory modalities, whereby one is "struck" by "sensible" and "affective" qualities alike, without really being able to distinguish "sensations" and "feelings". Yet the trigger-experience model of perception accounts for the experience of colours and shapes only on the basis of the retinal input, thereby leaving the "felt" sweetness – the value quality - without a specific organ, and the "feeling" without an organic basis. This is the first dissimilarity between perception and emotion. Perception is associated with certain sensory modalities and their specific organs, emotion is not. So, if the red tomato is not there to cause my visual experience of it, I only seem to see a red tomato (i.e., the experience is an illusion, being "just subjective"). On the contrary, not being intersubjectively shared does not make the emotion illusory: I may genuinely feel afraid of a barking dog, even if it is not in fact dangerous.
All subsequent objections of this sort emphasize the relative objectivity (or more or less universal accessibility) of (stimulus-based) perceptual information in contrast to the subjectivity of emotional responses.

1) Intersubjectivity/Idiosyncrasy: Unlike perceptions, emotions are highly dependent on the subject's motivations, beliefs, and character traits. You fear the barking dog, I don't.
2) Transparency/Self-affection: Unlike perceptions, emotions are not "transparent." They seem more to induce qualitative states in the experiencing subject, than to reveal qualities (e.g., colour, shape) of their object.
3) Cognitive Grounding: Emotions stand in need of justification ("Why are you indignant?" "What are you afraid of?"), unlike perception, which serves as a source of justification for judgements (Why do you say he is a thief?" "I saw him steal your necklace"). Emotions always seem to need a cognitive basis (perceptions, memories, beliefs, imagined situations),whereas perception provides such a basis for other states.

2. Pure feeling, or the two poles of feeling intentionality

Let's start with yet another terminological remark, concerning the word "emotion". Classic and contemporary philosophical literature alike tend to lump together a number of different phenomena related to feeling under this word, yielding a list where one can find toothaches or the pleasures of the table, along with the sin of pride or the passion for truth, a diversity that gives rise to what might be called the hodgepodge problem (De Monticelli 2003-2012). In his influential book on emotions, for example, Peter Goldie seems to speak of emotion in at least two ways: in a very broad and comprehensive way – as synonymous with "emotional feeling" (Goldie 2000: 51), including bodily feelings, moods, sentiments, and passions[2] – and in a narrower and more negative way, by contrast with moods and bodily feelings. That Goldie uses the word "emotion" in so broad a sense is also apparent in the general criterion he provides to distinguish emotions from "brute feelings like toothache, which we cannot make sense of" (Goldie 2000: 20). Emotions, accordingly, are characterized by their meaningfulness, or intelligibility, which is explained in terms of intentionality or directedness-toward-an object. This criterion is, by phenomenological standards, acceptable intensionally, yet too restrictive extensionally. To illustrate this extensional restrictiveness, consider how an emotion can

2 A construal strikingly similar to what Descartes meant by "passions", or the sort of conception of emotion found within the classical philosophical tradition

serve as a way of becoming aware of part of one's body, or of one's fragility or poor health. Even a toothache can make sense in that way, and so also can a physical pleasure, a global bodily feeling, a mood, and the like.

Now, a theory of the emotional life of a person should provide a rationale for distinguishing, ordering and connecting such different, yet somehow related, phenomena as those heaped together in the conceptual hodgepodge one finds in the literature. That is, such a theory has the task of linking together a variety of emotional phenomena that include local sensory feelings, like pleasure and pain in every sensory modality; global bodily states, like feeling ill or good, being tired, hungry, disgusted; moods and states of mind, like sadness, anxiety, cheerfulness; emotions of different kinds and levels (distinguishing emotions as basic, like fear, anger, disgust; non-basic, as shame, regret, guilt, indignation; and, more generally as social, moral, aesthetic, political, religious); personal feelings or sentiments (such as love and hate, esteem, respect or contempt, devotion), passions (like jealousy, gambling addiction, passion for truth), habitual but changeable dispositions/attitudes (e.g., trust or mistrust, sympathy, reliance on others), attitudes of personality (self-esteem, confidence, humility, curiosity), and so on.

Drawing on classic and contemporary research, a phenomenological theory of emotional life capable of providing a rationale for such a taxonomy and a map of the place of each type of affective phenomenon in a person's emotional life can be worked out on the ground of three main principles, which I've elaborated and defended elsewhere. (De Monticelli 2003-2012, De Monticelli 2016). I reproduce them here only in order to focus on their implications, which can clarify both the sense and the limits of the Perceptual Model of feeling.

The first principle is a version of the Direct Access Thesis (P), which I shall defend here by appeal to the just-mentioned general phenomenological theory of affectivity:

P1. Feeling is essentially a perception of the value-qualities, whether positive or negative, of things.

P2. All affective phenomena are "founded" on feeling (but not reducible to it).

P3. The feeling disposition, or affective sensibility, has a structure of layers ("stratification"), which are activated by felt values of a corresponding hierarchy of value-spheres, and activate corresponding levels of self-experience.

2.1. Bracketing drives

Claims P2 and P3 go against the grain of the current literature, though I contend that they alone do justice to the complexity of emotional phenomena. To be more specific, P2 clarifies the relation that holds among all affective states and dispositions in the taxonomy outlined above and identifies their common core, i.e., their feeling core, which, in Husserlian terms, is the founding (*Fundierung*) basis of all emotional experience. P2 articulates two "moments" characteristic of most emotional phenomena:

(a) *Being affected by* or receptivity, "passivity", being "struck" or "impressed" by something, in short, the *receptive component* of an emotional episode, marking it as a kind of perception.

(b) *Being inclined to,* "moving" to or from, drives/desires, or in short the *conative component*, the urge to action.

The foundational relation is a relation of *ontological dependence*. That the receptive component provides the foundation for the conative one means, first of all, that the latter cannot exist without the former, but the former can exist without the latter. There are lots of examples of this second possibility, whereby a receptive experience can exist without a conative one. To illustrate, think of aesthetic experience, e.g., listening to music and discriminating its aesthetic and expressive qualities, or think of the way a mother lovingly contemplates her sleeping child, or, again, of states like blissfulness, calm, despair, surprise, amazement, and the like.

Conversely, we can hardly conceive of a drive, a conative urge, deprived of any feeling. So, far from being reducible to excitability, *emotional experience is based upon sensitivity*, and excitability, when the former is in effect, tends to be conditioned by it (rather than directly by the "stimulus").
This claim goes against a prominent tradition, according to which what is essential and ineliminable in affectivity is its conative component (in the form of, e.g., drives or desires). This tradition is exemplified by theorists like Freud, but can be traced back to Plato, Aristotle, Augustine, Thomas Aquinas, Descartes, and Hume, among others. Even Franz Brentano is still completely faithful to this tradition and, as a consequence, does not divide affective and conative states into two distinct classes but rather fits them all into one and the same class[3]. Hence, P2 presents a revolutionary challenge in philosophy of mind and personhood – a genuine discovery on the part of phenomenology, and a motivation for its cognitivism in axiology. As Kevin Mulligan aptly remarks:

> The view that it is possible to feel values, what I have here called being struck by value, and that this is a form of epistemic contact with value was defended by a quintet of realist phenomenologists – Reinach, Scheler, Hildebrand, Hartmann and Reiner (Mulligan 2009: 154).

According to this theory, e*motions* are *not* perceptions of values: they are *personal responses and re-actions* to such perceptions that are, as it were, built on them. P2 allows one to "bracket" the drive-component of affectivity, identifying it as distinct but dependent on the latter and thus breaking the ground for a thorough analysis of "pure feeling", that is, of its *peculiar* intentionality. Within this frame, in fact, P1 and P3 shed light, respectively, on *the objective and subjective poles* of the intentional relation – thereby explaining what is peculiar about any mode of consciousness involving *feeling*.

2.2. Originally giving consciousness and positionality

What is characteristic of feeling, peculiar to it as a form of intentionality? Let's frame the answer within a more general understanding of intentionality. As is well known, this term denotes the key-notion for the analysis of consciousness. There is no consciousness but in a given mode – whether as perception, emotion, desire, belief etc. Now a genuine phenomenology of consciousness consists in distinguishing the way in which objects are present to a subject in accordance with modes such as these. There is no reason to doubt that Husserl accepts P1 as a basic axiological principle and does so quite early on in his philosophical labors. Yet no classical phenomenologist brought out the importance of *both* poles – subjective and objective – of the intentional relation as well as Husserl. Let's go back to the notion of *Erfahrung*, as clarified above. Not every mode of consciousness is an actual encounter with – let alone an exploration of or interaction with –reality. The paradigmatic form of *Erfahrung* is external perception. And, as we saw already, Husserl has a technical term for this "epistemic contact" involving some mode of presence of a perceptual object. He calls it an "originally giving" consciousness, or a presence of the object in the flesh. It is a mode of presence requiring further analysis on the "noematic" side, i.e., on the side of the object. We made mention already of that peculiar feature of the *Seinsinn* (the "noema") of an object of perception by virtue of which it is given as further explorable, indeed, even as an

3 According to Brentano there are three classes of mental states: Presentations (*Vorstellungen*), Judgements (*Urteile*), Interests or Emotions (*Interesse und Gemutsbewegungen*) – *Ideae, Judicii, Volitiones* in the Latin version (Brentano 1973: 198).

infinite source of further experience. There is much more than *intuitive presence* to "original" presence. To see why, try to count the pillars of the Parthenon just by recalling it to your memory. Your memory makes possible a re-visualizing of things, and so in attempting this feat you will have an intuitive mode of presence of the Parthenon. But you will not be in a position to increase your knowledge of the object. Unless you already knew the number of the temple's pillars, you won't be able to (without imaginative embellishment) count them on your imaged temple.

One might think that this difference is captured by a standard (Searlian) analysis of the mental state's content in terms of direction of fit plus causal self-referentiality (enjoyed by perception, but not by memory). But how could *causal* self-referentiality grant the possibility of illusion and correction, "learning from experience", which is *constitutive* of perception? Illusion and correction would be impossible if experience were only a causal impact of reality on sensory organs. There is a *claim of validity* (veridicality) to any genuine perception, that no sheer causal impact of things on organs can produce. What "bears" the implicit claim of validity is a feature of perception which makes an act out of it – rather than a state (De Monticelli 2015). We "take notice" of the (apparent) existence of the perceived object, or "take a position" concerning it. The *doxic positionality* of an intentional act is the noetic correlate of the givenness and further explorability of the act's object, and exists quite independent of any propositional attitude, thought or conceptually articulated belief. It consists, rather, in the *recognition* of the given thing as existing or being there. We "take note" of something being there or happening, and this is not at all subject to the will. That's why radical doubt is impossible when it comes to actual perceiving. This *Stellungnahme,* in which the person "filling in" the subjective pole of intentionality "lives" while experiencing, is subject to the "jurisdiction of reason", though. That is, it can turn out to be *wrong,* and is subject to being "cancelled" for a new position, or suspended in hesitation, uncertainty, further exploring – all of which takes place at a pre-linguistic level (the level of the "pre-reflective cogito", as Merleau Ponty would have it). Types of experience like imagination or day-dreaming, by contrast, present us with contents whose veridicality is not an issue. In them positionality is "neutral," so that there is no claim of "validity" at this level. I'm perfectly free, for instance, to imagining flying donkeys.

This two-sided "noematic-noetic" analysis of perceptual intentionality is the background against which an analysis of feeling intentionality can be now outlined.

3. A brief phenomenology of feeling

In a footnote to a working paper on emotional life, Husserl quotes William Hamilton, who

> "says that in feeling we are in a peculiar way one and the same with its content" (Husserl 2005: 165).

This and similar observations constitute a cornerstone for the phenomenology of feeling, and highlight a key point of contrast with the phenomenology of perceiving. The decisive move enabling the phenomenologist to adequately describe pure feeling consists in taking note of the intentionality of feeling both as a mode of presence of *objects* to a subject and as a mode of presence of the *subject* to herself. The two poles of the intentional relation are in fact the targets of P1 and P3, respectively.

3.1. The feeling dimension of breadth

"Noematic" description highlights our receptivity to an infinite variety of value qualities belonging to things, making "goods" and "evils" out of them. Receptivity is the phenomenon of "being struck" – wrongly or rightly –, in emotional responses which are more or less appropriate and in principle correctible (as illustrated in cases when, for instance, we come to regret our responses and the course of action based on them). Axiological positionality, which may be positive, negative or neutral, is as much "under the jurisdiction of reason" as doxic positionality. Up to this point the analogy with perception is fitting, and to this extent the "perceptual model" of feeling is justified.

I may feel the unpleasantness of a sting, the bodily or psychological discomfort associated with a state of illness or weariness, the agreeable nature of an arrangement of colors. But I may also sense the nobility of a gesture, the vulgarity of an attitude, the wickedness of an act, the beauty of a masterpiece. The harmonious way a tool or a piece of furniture fits one's body, the easy form of a teapot – these and many others are examples of the sort of "affordances" an object can have. These qualities too are in a way "perceived," and feeling is the mode of this kind of perception. Emotion is in that respect essentially *like* perception. It is, as the German language has it, a *Wertnehmen.*

To summarize, feeling is the *mode of presence* of things' value qualities (saliences, affordances in Gibson's sense[4], or tertiary qualities of all sorts). This point addresses the question of the formal object of emotional intentionality. It points to a widely neglected, yet pervasive feature of the life world, namely, the rich and varied pluralityof positive or negative value qualities that "colour" things, events, states of affairs, and situations in the surrounding world. Indeed, it is hard to find qualifying words, adjectives, that do not refer to some value quality. Axiology is in a sense the ontology of adjectival language.

Yet "noetic" description has to take into account the other feature of feeling, the one noticed by Hamilton and explored by the classic phenomenologists. The clearest noetic account is given by Edith Stein, both in her (1917) text on *The Problem Empathy* and in her (1922) on the *Philosophy of Psychology and the Humanities.* Phenomenologically, experience (in the sense of *Erfahrung*) is always also self-experience, and not in the trivial sense of *Erlebnis*, but in the strict one of a discovery of oneself, or self-cognition. Yet this happens *only to the extent to which experience includes feeling and action or movement.* I can be "lost in perception" or in thought, completely oblivious of myself. But I cannot feel pain without feeling myself in pain, nor can I act without experiencing my agency or my causal power – except in some specific "de-personalizing" psycho-pathologies[5].

We can call "egological states" those experiences that belong to the affective and the conative classes of intentional acts. Feeling states, in particular, are egological states, inasmuch as they are both experiences of the (axiological aspects of) the world and self-experiences. "Being struck" is not all that there is to pure feeling. It may also be described as an experience of being touched, or "affected", "involved", "impressed," or even awakened, "revived", excited, etc. Such characteristics make up the noetic counterpart of affective sensibility. Only in feeling do we feel alive. Further, what is deeply interesting to a philosopher is the extent to which feeling – and hence self-experience – is involved in intellectual research. Epistemic values and disvalues are a source of deep suffering or joy for some of us. A phenomenology of boredom,

4 J. J. Gibson (1966): "I have coined this word as a substitute for *values,* a term which carries an old burden of philosophical meaning. I mean simply what things furnish, for good or ill. What they *afford* the observer, after all, depends on their properties". Quoted in Jones (2003: 111).

5 Take, for example, the case of compulsion described by Viktor Emil von Gebsattel (1994), where poor Mr H. H. seems to have entirely lost his sense of agency – a predicament that compels him to desperately repeat each action and segment of action he has to accomplish.

too, could be worked out along these lines. Dullness and estrangement seem intimately bound up with indifferent activities.

3.2. The depth dimension of feeling

Phenomenological analysis reveals a further dimension of feeling's intentionality, which we can call its *depth*. Thus far we have considered one dimension, namely, the *breadth* of its domain, or horizon. The qualities of value - positive and negative to varying degrees - are *many*. The extraordinary *richness* of the negative or positive value qualities to which we are sensitive provides the starting point for any phenomenological reflection on feeling. There is virtually no situation in life in which one of these qualities is not currently or habitually present.
Feeling, moreover, is always a self-revealing experience. Value-experience is always self-experience as well, to the extent that it "touches," involves, or concerns us. Doubtless, not all feeling experiences are on a par. There are differences in importance or weight, or, we might say, in rank of value, or in motivating power – of the felt goods and evils. – This in a way corresponds to the "depth" of the feeling experience, say, of the pleasure or the pain of waking oneself up. As if values of different rank were felt at different levels of oneself (or as involving a lesser or greater part of oneself). It is as if the experience of different values belonged to different layers of sensibility. That idea is described well by Scheler:

> There can be no doubt that the facts which are designated in such a finely differentiated language as German by "bliss," ''blissfulness" [*Glückseligkeit*], "being happy" [*Glücklichsein*], "serenity," [*Heiterkeit*], "cheerfulness" [*Fröhlichkeit*], and feelings of "comfort" [*Wohlgefühl*] ," "pleasure," and "agreeableness"[*sinnliche Lust und Annehmlichkeit*] are not simply similar types of emotional facts which differ only in terms of their intensities..." (Scheler 1973: 330)

Intuitively, we realize that a feeling can touch a person more or less "deeply", depending on *the degree of personal involvement*. For instance, the pleasure of a good massage can be felt by most of us as much less involving than the joy of discovering Shakespeare. No doubt this joy will have *a higher degree of motivational power* than the pleasure of the massage, and it might indeed motivate a choice to study English Literature rather than something else, a turn of events with significant consequences for the rest of my professional life.
Is it possible to give, if not a way to measure depth, at least a rationale for the putative ordering of layers of sensibility concerned, respectively, with the pleasure of a massage and the joy of reading Shakespeare? What is the rationale for this stratification?
A last passage from Scheler offers a powerful suggestion:

> It is, for example, impossible for one to be ''blissful" over happenings *of the same axiological level* that are "disagreeable" to another; the differences in these feelings also seem somehow to require *different axiological states of affairs*. (Scheler 1973: 331, emphasis added)

Let's unpack Scheler's remark. The "depth" of a feeling is proportional to the importance of the values concerned. Feelings are modes of presence of values *at different levels of an axiological hierarchy*.
Many scholars and philosophers are dubious of the sort of objective value hierarchy that Scheler offers in his *Formalism,* with its four spheres of value, i.e., the sensory, the vital, the personal and the divine. Yet it's debatable whether they have understood its phenomenological grounding. Everyday experience and ordinary language already point in the direction of this view as providing a criterion for ordering the hodgepodge of affective

phenomena, the very thing we are looking for. This criterion, in turn, is intimately bound up with our selfhood, as experienced through all the different forms of feeling.
By "selfhood" I mean the reality of a person *as she experiences herself, or as she is "given to herself",* in that quite peculiar and irreducible way in which she is not given to anybody else, that is, from a first person perspective (Baker 2013). Any value experience is at the same time an experience of selfhood, though in varying degrees of centrality, or of wholeness, so to speak. This seems phenomenologically undisputable. The relatively localized pleasure of a massage is an experience of those parts of my body, which are touched, as *mine*, or as parts of my (bodily) self. The pain I feel in when you step on it teaches me the boundaries of my bodily being. Similarly, any global, "vital" comfort or discomfort, hunger, drowsiness, feeling refreshed or tired, the energetic feeling of young age or that of powerlessness in old age – all of these inform me of *how I am,* what condition I am in, as well as what I need insofar as I am in a constant state of vital self-regulation. They also inform me of my finitude, my dependency on the material and social environment. My global state is exactly what basic emotions alter, by flagging up threats and advantages in the surrounding world. Moods as well are indicative, both to myself and others, of *how I am*, signalling which of the affective tones, in a scale ranging from deep depression to manic euphoria, now colours my life.
Only encounters with what kindles sentiments in me - whether of admiration and contempt, love and hate, political passions, religious or artistic devotions - will, over time, tell me *who I am.* By *sentiments* I mean those relatively abiding dispositions to assent to or dissent from the very being of someone or something. The peculiarity of sentiments is to present their objects to us as bearers of yet unknown values, or as fields of axiological discovery. They instigate a kind of search as exciting and fallible as any other search after truth, and mark out a significant domain of the life of a person. They nourish and motivate higher-order emotions, emotions which would be impossible in the absence of their sentimental background, but also choices, actions, and behaviours. The kind of hope I have for a friend's success would be unknown to me without the friendship binding me to him, and neither would the pleasure of benevolence and other friendly behaviours towards him. In so doing, sentiments gradually reveal our personal identity – or rather a part of it, whether it be an essential or contingent one. In the course of our personal growth – and humans, as long as they live, in some respects never cease to grow up – our sentimental dispositions lead us to discover, wonder about, change, establish or disrupt *our* individual scales of value priorities and our systems of value preferences. Sentimental life, in this expansive sense that *grounds* our long-term desires, passions, intentions, choices and actions, is what first awakens what we may call the personal layer of our sensibility, which will in a way in return take root in the subsoil of evaluative feeling and acting at the level of its more basic, sensory and vital layers. Sentiment is always indicative of the growing core of selfhood, or personal identity, whose discovered aspects are often astounding or disconcerting to us, in exactly the same way as newly gained knowledge of the real can be in other cases.
Let's think again of that initial list of affective phenomena in the familiar hodgepodge of emotional experiences. We can now discern a clear order to them, the rationale for this order being their place and role in the sort of self-experience we go through by exploring the axiological aspect of reality, from sensory pleasures and pains to global bodily feelings and moods, and on to basic emotions, sentiments structuring our axiological scales, including higher emotions, will-shaping passions, and even personal callings. (P3) is the principle summarizing all the preceding analysis.

4. The objections reconsidered

What is left of the three objections to the Perceptual Model of feeling that we considered above (§1)? Let's recall them:

1) Intersubjectivity/Idiosincraticity: Unlike perceptions, emotions are highly dependent on the subject's motivations, beliefs, and character traits. You fear the barking dog, I don't.
2) Transparency/Self-affection: Unlike perceptions, emotions are not "transparent." They apparently modify their subject, inducing qualitative states in her/him, and do not simply reveal qualities (colour, shape) of their object.
3) Epistemic Grounding: Unlike perceptions, emotions are in need of justification ("Why are you indignant?" "What are you afraid of?") and do not provide a source of justification for judgements (Why do you say he is a thief?" "I saw him steal your necklace"). They always stand in need of a epistemic support (*via* perceptions, memories, beliefs, imagined situations) and do not provide epistemic support for other states.

Our analysis shows:
a) that the objections hit the mark in so far as the analogies between *emotions* (in the proper sense) and perceptions are concerned;
b) that, nevertheless, they lose their force when it comes to the feeling core of emotions and its receptive core;
c) that the source of the differences is also what grounds the analogy with perception, namely the egological character of feeling: being struck by value is a sort of perception, but experiencing values is not separable from experiencing selfhood. This is why feeling has such a central role in our lives.

Examining the objections in greater detail will help us see these points. Let's start from the second one. It finds warrant in the same fact that struck Hamilton and innumerable other authors. In phenomenological terms, feeling is "egological" or self-constitutive (as is well known, "constitution" does not mean construction or creation, but intuitive access to the evidence-filling content (*erfüllende Inhalt*) of a given notion). Feeling is, in this sense, also an essential component of any minimal bodily self-awareness as it figures in, for instance, sensory modalities such as the sense of touch, kinesthesia (sense of agency, or of posture and spontaneous movement) and cenesthesia (visceral feelings) a point that Husserl's analyses of mind-and-personhood's constitution (*Ideas* II) and Merleau Ponty's *Phenomenology of Perception* famously demonstrated.
Objection 2 (Transparency/Self-affection) refers to a phenomenon thoroughly studied by classical phenomenology. Drawing on their legacy, we can make sense of the disanalogy between perception and emotion. Both sensory and affective sensibility are modes of experience, in the strong sense of *Erfahrung*. They both exhibit the mode of presence of an "originally giving" consciousness. We also characterize this mode as immediacy or directedness. For both sensory and affective sensibility, it is a very common mistake to confuse immediacy with instantaneity. Experience, strictly speaking, always takes time. External perception in all its modalities is intimately tied up with action and exploration. The way in which feeling is temporally extended is somewhat different. Just as the being of any given visible object has much more to itself than what meets the eye at the first glance, likewise value– the axiological quality of a situation, a thing, a person, a work –always has much more content than the "aspect" of it that is sensed at first. But gaining further knowledge, in this case, is not so much a matter of further *exploring* (i.e., an active testing of some kind) as it is a matter of *contemplation*, so to speak. We can take aesthetic experience as a paradigm of value experience here. Notice that the first effect of being struck by aesthetic value is being halted

and brought to a standstill, as it were. We tend to take up the attitude of the spectator, a sort of contemplative receptiveness. Consider the glare of the sea on a midsummer late afternoon. We see it, but if we have time and leisure to contemplate it, we may gradually be somehow "filled" with a kind of impersonal bliss, up to a point where we feel almost overwhelmed. Perhaps we search for a word to name this blissful effect of the glaring sea, maybe an ancient Greek verse comes to mind, one from our school memories. "Innumerable smile of the sea", that's how Aeschylus names this bliss, capturing that mixture of calm infinitude and the vibrant plurality of glimmers which make it up, and reminding us of the blissful life of the gods. Well, the forgoing is intended as an illustration of a central phenomenological feature of feeling, i.e., that it takes time. And it does so in a double sense: it requires time, and it fills it, conferring *plenitude* to it. It calms us down and makes us silent. We feel ourselves alive, more intensely than in our ordinary practical commerce with things. It is as if all our time feeling were a time of growth. It is a time of growth indeed. What increases is our awareness of value (or disvalue, of course). –True, most of the time we are content with that value which first "strikes" us and go on with our busy lives. But if we let ourselves be "invested" with the positive or negative appeal of the thing, then –quite slowly, sometimes – we come to "realize" the nature of that value, becoming increasingly aware of its peculiar "thickness", while it in a way "permeates" us, so that "we are in a peculiar way one and the same with its content," as Hamilton has it.

Among the classical phenomenologists, Moritz Geiger is the one who most clearly articulated the difference between perceiving and feeling . He used the German verb "*innewerden*" – literally, to be internalized – to denote that process of becoming aware of some value or disvalue through that inner reawakening we described (for further illustration, I defer here to Scheler's account of the experience of repentance, to which he dedicated a famous essay in 1921). There is a German word, *Widerfahrnis* that is particularly well suited to express this sort of receptive encounter with some value that "sails into us", to use the translation proposed by Herbert Spiegelberg (1982: 326). The term aptly describes the experience that a discerning listener has in hearing music. That is to say that this mode of presence of a subject in feeling, far from preventing an encounter with something outside the subject, *is* a mode of original presence of value – a typically non-objectifying mode. There is a very apt phrase for describing this attitude, which Geiger eminently applies to aesthetic experience, but which is nevertheless just as apt for value experience generally. He describes it as "external concentration".

The classic phenomenologist who analyzed the degrees of sentimental intuitive knowledge with the greatest precision is Dietrich von Hildebrand, who introduced an important distinction concerning the experience of value. On the one hand, there is the actual perception of value – "*Wertsehen,*" the experience of its first coming into view, so to speak, or, of noticing an initial aspect of that value. On the other hand, there is the process of being genuinely, more or less deeply affected by it ("*Wertfühlen*"). And this is certainly not the last noteworthy feature of subjective our involvement in value cognition. There is, further, always some degree of *familiarity*, and perhaps even a habitual expertise relative to some class of values (think of a musician, or of a sommelier).

Going back to the second objection (Transparency/Self-affection), we can now concede that it is right, in a sense, although, as things stand, more needs to be said about why it is right. Its truth is much more evident when the feeling core of the emotion is separable from any drive, as in aesthetic, moral or religious emotion, in contrast to the emotions of the practical life (such as anger or fear, which are much more object- than self-aware). But in that extent, it is right for different reasons than the ones that its proponents would likely appeal to. The reason is not because of a lack of a *direct* intentional object in emotions, but because of the direct object's *mode of presence* in the case of feeling consciousness. Moreover, an authentically

experienced affection of the subject takes place at a deeper level if the feeling in question has features opposite to those of the so-called "basic" emotions, which are understood in recent work in psychology as "affect programs," characterized by automatism, quickness and cognitive impenetrability (Deonna and Teroni 2012: 20). Think of the difference between experiences like bursting into tears of anger and frustration, as young children often do, and that of being moved to tears of gratitude and repentance in realizing how much more we have been given by those who love us than we deserve . Think of how simplistic and short-sighted the listing of all emotional phenomena on the "fast thinking" side of Kahneman's two systems theory (Kahneman 2011) would be. And yet, a tendency to associate the emotional with the "fast" and "automatic" side of mental life is quite common today.

Growth of all sorts is slow, and a person's growth in self-awareness is no exception. Self-awareness is awareness of what is more or less important for us, i.e., the values priorities (lived as objective) and the value preferences (lived as subjective) that define us, constituting our moral and our personal identity. This leads us to the other two objections. The first one (Intersubjectivity/Idiosincraticity) is only moderately persuasive. First of all, the things we spontaneously *perceive* in the real world are strikingly different pieces and aspects of it, depending on the subject's motivations, beliefs, and character traits. Secondly, differences in emotional responses also concern the conative component of emotions, which, indeed, are absent in perception as such. Thirdly, as far as the feeling core of emotion is concerned, different people can in fact grasp different aspects of value in things, and possibly come to be more deeply affected by them. Such differences depend on the axiological structure of each one's sensibility or selfhood. So there is indeed a point of divergence between emotion and perception. Yet the objection, as it stands, does not shed light on what accounts for this difference.

As to the third and last objection to be considered (concerning the epistemic role of perceptions/emotions and the need for justification), I respond, first of all, by maintaining that axiological positions, in the basic cases, do not differ from doxic positions in their relation to the will. Whether or not I am afraid of an aggressive dog or disgusted by a rotten corpse is no more in the power of my will than whether or not I recognize the real existence of the chair I stumble upon. Secondly, the non-arbitrariness of positions, in both cases, does not prevent them from being subject to perplexity or illusion, and therefore potentially standing in need of justification. Perhaps we tend to get visible facts right more often than we get their value-aspects right, but our perceptual certitudes are subject to error just as much as our emotional responses are. As far as the interesting part of the objection is concerned, it is true that we are sometimes shocked by the diversity in people's values scales, depending on cultures and individual personalities. Yet, in principle, we have to attribute equal moral competence to all, and hence we must also attribute sensibility to universal values (and value priorities), to persons with different personalities and cultural backgrounds. Otherwise we must give up moral universalism, as well as the principle of equality of all humans in dignity and rights. That is the challenge any theory of feeling must confront. The ultimate meaning of the Perceptual Model of feeling should be assessed against the background of this challenge.

REFERENCES

Bagnoli, C. (ed.) (2011-2013), *Morality and the Emotions*, Oxford University Press, Oxford;
Baker, L. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
Baine, T., and Montague, M. (2011), *Cognitive Phenomenology*, Oxford University Press, Oxford;
D'Arms, J., and Jacobson, D. (2000), "Sentiment and Value", *Ethics,* Vol. 110, pp. 722–748;
Deonna, J. A., and Teroni, F. (2012), *The Emotions – A Philosophical Introduction*, Routledge, London/New York;
De Monticelli, R. (2003-2012), *L'ordine del cuore – Etica e teoria del sentire*, Garzanti, Milano;

De Monticelli, R. (2015), "Experience and Value - A Phenomenological Approach to the Dilemma of Normative Supervenience"*, Metodo - International Studies in Phenomenology and Philosophy*, Vol. 3, No. 2, pp. 195-214;
De Monticelli, R. (2015), "Acts, Normativity, Personhood - Outline of a Theory of Embodied Rationality"*, Les Cahiers Philosophiques de Strasbourg*, 38, pp. 175-194;
De Monticelli R. (2016), "Sensibility and Values. Toward a Phenomenological Theory of the Emotional Life", In Rinofner-Kreidl, S., and Wiltsche, H. A. (eds.), *Analytic and Continental Philosophy - Methods and Perspectives*. Proceedings of the 37th International Wittgenstein Symposium, Series: Publications of the Austrian Ludwig Wittgenstein Society – New Series (N.S.), 23, pp. 381-400;
Dokič, J., and Leymarie, S. (2013), "Are Emotions Perceptions of Value?" *Canadian Journal of Philosophy*, Vol. 43, No. 2, pp. 227-247;
Gebsattel von, V. E. (1954), "Die Welt des Zwangkranken", In Id., *Prolegomena einer medizinischen Anthropologie*, tr. it. in *Antropologia e psicopatologia*, Bompiani, Milano, 1967, pp. 59-168;
Geiger, M. (1921), "Fragment über den Begriff des Unbewussten und die psychische Realität", *Jahrbuch für Philosophie und phänomenologische Forschung*, 4, pp. 1-137;
Goldie, P. (2000), *The Emotions: A Philosophical Exploration*, Oxford University Press, Oxford;
Goldie, P. (2004), "Emotion, Feeling, and Knowledge", In Solomon R. C. (ed.), *Thinking about Feeling*, Oxford University Press, Oxford, pp. 91–106;
Hartmann, N. (1926), *Ethik*, 3 Bänden, De Gruyter, Berlin, 1962;
Hildebrand, D. (1922), "Sittlichkeit und Ethische Werterkenntnis", *Jahrbuch für Philosophie und phaenomenologische Forschung,* 5, pp. 462-602;
Husserl, E. (1900-1901), *Logische Untersuchungen,* Zweiter Teil, *Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*, In zwei Bänden [*Logical investigations*, Second part, *Investigations concerning phenomenology and the theory of knowledge,* In two volumes], Panzer, U. (ed.), Max Niemeyer, Halle, 1901; rev. ed. 1922; HUA XIX, Martinus Nijhoff, The Hague, 1984;
Husserl, E. (1994), *Briefwechsel. 10 Vols*., Schuhmann, K., and Schuhmann, E. (eds.), Kluwer, Dordrecht, Vol. VII;
Husserl, E. (2005), *Wahrnehmung und Aufmerksamkeit, Texte aus dem Nachlass* (1893-1912), Vongehr, T., and Giuliani, R. (eds.), HUA Bd XXXVIII, Springer Verlag, Berlin;
Husserl, E. (1952), *Ideen zur einer reinen Phänomenologie und phänomenologischen Philosophie,* Zweites Buch: *Phänomenologische Untersuchungen zur Konstitution,* Biemel, M. (ed.), HUA Bd IV, Martinus Nijhoff, Den Haag;
Kahneman, D. (2011), *Thinking, Fast and Slow*, Farrars Straus and Giroux, New York;
Jones, K. S. (2003), "What is an affordance?", *Ecological Psychology*, Vol. 15, No. 2, pp. 107-114;
McDowell, J. (1985), "Values and Secondary Qualities", Reprinted in Id. (1998), *Mind, Value and Reality*, Harvard University Press, Cambridge, MA, pp. 131–150;
Merleau-Ponty, M. (1945), *Phénoménologie de la perception*, Éditions Gallimard, Paris;
Mulligan, K. (1998), "From Appropriate Emotions to Values", *The Monist,* Vol. 81, No. 1, pp. 161–188;
Mulligan, K. (2009), "On Being Struck by Value: Exclamations, Motivations and Vocations", In Merker, B. (ed.), *Leben mit Gefülen – Emotionen, Werte und ihrer Kritik*, mentis, Paderborn;
Noë, A. (2015), *Strange Tools – Art and Human Nature*, Hill and Wang, New York;
Reinach A. (1989), *Sämtliche Werke*, Textkritische Aufgabe in Drei Baende, Schumann, K., and Smith, B. (eds.), Philosophia-Verlag, München;
Scheler, M. (1916), *Der Formalismus in der Ethik und die materiale Wertethik*, Francke Verlag, Bern, 1980; *Formalism in Ethics and Non-Formal Ethics of Values*, Engl. Transl. Frings, M., and Funk, R. L. (transl.), North Western University Press, Evanston, 1973;
Scheler, M. (1921), *Reue und Wiedergeburt*, in *Das Ewige im Menschen*, Gesammelte Werke V, Francke-Verlag, Bern/München, 1954-2000;

Spiegelberg, H. (1982), *The Phenomenological Movement*, Martinus Nijhoff, The Hague;
Stein, E. (1917), *Zum Problem der Einfühlung*, Reprint Kaffke, München, 1980, Neuausgabe in Edith-Stein-Gesamtausgabe, Bd. 5, Herder, Freiburg, 2008;
Stein, E. (1922), "Beiträge zur philosophischen Begründung der Psychologie und der Geisteswissenschaften", *Jahrbuch für Philosophie und phaenomenologische Forschung*, 5, pp. 1-284, English Transl. Baseheart, M. C., and Sawicki, M. (transl.), *Philosophy of Psychology and the Humanities, The Complete Works of Edith Stein,* Vol. 7, ICS Publications, Washington, 2000;
Tappolet, C. (2000), *Emotions et Valeurs*, PUF, Paris;
Tappolet, C. (2011), "Values and Emotions: Neo-Sentimentalist's Prospect", In Bagnoli, C. (ed.), *Morality and the Emotions*, Oxford University Press, Oxford, pp. 117–134.

CECILEA MUN
Independent Scholar
cecileamun@icloud.com

THE RATIONALITIES OF EMOTION[1]

abstract

I argue that emotions are not only rational in-themselves, strictly speaking, but they are also instrumentally rational, epistemically rational, and evaluatively rational. I begin with a discussion of what it means for emotions to be rational or irrational in-themselves, which includes the derivation of a criterion for the ontological rationality of emotions (COR^e): For emotion or an emotion there exists some normative standard that is given by what emotion or an emotion is against which our emotional responses can be judged or evaluated in virtue of the fact that our emotions manifest our rationality. I conclude with a brief discussion of the implications of this account.

keywords

emotions, rationality, irrationality, knowledge, epistemology

Phenomenology and Mind, n. 11 - 2016, pp. 48-57
DOI: 10.13128/Phe_Mi-20105
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

According to one entry in the *Oxford English Dictionary,* the word 'rational' is to be understood as a noun that refers to the following:

> The rational part of the human mind; the power or faculty of reason; that which is rational or reasonable; a conjunction that expresses a reason for a statement; a rational being; *spec.* a human being, as opposed to an animal[2].

A second entry defines the word 'rational' as an adjective or adverb, suggesting the following of those things that are rational:

> Having the faculty of reasoning; endowed with reason. Esp. in rational being, rational creature, rational soul, etc.; that uses, or is capable of using, the faculty of reasoning; having sound judgement; (in extended use) sensible, sane, lucid. Also: characterized by reasoning, as opposed to emotion, intuition, etc.; of or relating to reason or reasoning; that exists only in the mind; having no actual physical presence; based on or derived from reason or reasoning, esp. as opposed to emotion, intuition, instinct, etc.; involving or characterized by any of various methods of analysis or planning based on the calculation of a projected result; designating such a method; in accordance with reason; reasonable, sensible; not foolish, absurd, or extreme; of a conjunction: that expresses a reason for a statement; and in a rational manner; rationally[3].

The first entry suggests that what is rational is closely associated with the faculty of reasoning, and may be applicable only to human beings as a species. The second entry adds to the first

1 I would like to thank the members of the Southern Society of Philosophy and Psychology, the San Raffael Spring School of Philosophy, Cheshire Calhoun, Lani Shiota, Bernard Kobes, and the editorial staff with *Phenomenology and Mind*. All your comments, questions, and suggestions helped contribute to the quality of this paper. Finally, I would also like to thank Arizona State University, Clemson University, St. Mary's College of Maryland, and GateWay Community College for your support.

2 "Rational, n.1." OED Online. December 2015. Oxford University Press. http://www.oed.com/view/Entry/270902?rskey=7eJTu9&result=1 (accessed December 16, 2015).

3 "Rational, adj. and adv." OED Online. December 2015. Oxford University Press. http://www.oed.com/view/Entry/158502?rskey=7eJTu9&result=3 (accessed December 16, 2015).

by indicating a relationship between what is rational and what is a sound judgment, sensible, sane, lucid, not foolish, not absurd, and not extreme, while also illustrating the highly mental or intellectual nature of what is rational by contrasting it with what is physical, emotion, intuition, and instinct. Furthermore, the first entry's emphasis on human beings, along with the second entry's contrast between what is rational and what is physical, emotion, intuition, and instinct imply that both human infants, some people with disabilities, and non-human animals lack the capacity to be rational[4]. Finally, both entries suggest a close relationship between the notion of rationality and reasons as justification for holding beliefs or claims of knowledge.

Consistent with the above account of what is rational, some philosophers have argued that it makes no sense to speak of emotions as being rational *in-themselves*, i.e., that *emotions qua emotion* or *qua emotion-types* are neither rational nor irrational (Rorty 1985; Prinz 1994)[5]. Rorty's observation makes two important points (Rorty 1985: 351): First, it suggests that emotions qua emotion or qua emotion-types are neither rational nor irrational in-themselves. Second, it suggests that considerations about the rationality of emotions themselves are considerations about the rationality of an individual in light of the fact that an individual is having an emotional experience. It is the rationality of the individual that is under consideration and not, strictly speaking, the rationality of emotions.

I accept the second point, but argue in §2 that emotions *are* at times rational and at times irrational in-themselves, qua emotion or emotion-types, and in §6 I argue that our emotional responses are rational in that they rely on a faculty of reasoning; that they may be analogous to sound judgments; that they may be reasonable, sensible, sane, lucid; not foolish, absurd, or extreme; that they are types of reasoning; that they are based on or derived from reason or reasoning; that they are in accordance with some kind of reason; that they involve some method of analysis or planning based on the calculation of projected results; that they may express reasons for statements; that they can be experienced in a rational manner; and that they may be analogous to reasons as justifications for holding beliefs or knowing. In doing so, I deny that emotions are opposed to reason, that what is rational exists only in the mind and has no physical presence, and I open the door to a conceptual space that allows us to understand infants, some people with disabilities (who are thought of as otherwise), and non-human emotional animals as being rational and capable of knowing.

2. The Criterion for the Ontological Rationality of Emotions (COR^e)

When I speak of the *rationality* of emotions, I am referring to a *capacity* for being at times rational and at times irrational. When we speak of the rationality of an individual in respect to their emotional experiences—for example, when we judge one's anger to be rational or irrational—we are speaking of the individual's capacity to reason *in virtue* of their emotional experiences. Furthermore, RATIONAL is a comparative concept. To say that something is "rational" only makes sense if there is some comparative sense in which we can say that *that* thing is irrational and *vice versa*. RATIONAL is also a normative concept[6]. To speak of an individual's "rationality" or the "rationality of emotions" is to suggest that there is at least some norm against which an individual's rationality can be evaluated as being either rational or irrational. This normative standard is typically cashed-out in positive terms as defining what is rational, and what is irrational is typically placed in opposition to this normative standard.

4 John Deigh (1994) applies a similar line of reasoning in his analysis of cognitive theories of emotion.

5 Two relevant distinctions ought to be made here: 1) a distinction between types of emotion *de re* and *de dicto*, and 2) a distinction between types of emotion *de re*/*de dicto* and emotion as a *genus*. By 'emotion-types' I mean types of emotion *de re* rather than *de dicto*, which are to be distinguished from emotion as a *genus*.

6 A term put in all caps indicates a reference to a concept.

Finally, to say that emotions themselves are at times rational and at times irrational is not to deny that the individual is the one who is being rational or irrational in having an emotional experience. Theoretical discussions about what emotions are do not hold that emotions ought to be understood as entities in their own right, as if they existed independently of our (or any emoter's) mental or physiological capacities, despite the way that some theorists speak of emotions. Theories of emotion instead presuppose that emotions are essentially aspects of such mental or physiological capacities. When theorists speak of emotions, they are speaking of aspects of an individual. Thus to speak of the rationality of emotions in-themselves is to say that there exists some normative standard—given by what emotion or an emotion is—by which our emotional responses can be judged or evaluated in light of the fact that they manifest our rationality. If there are such standards for emotions, then emotions qua emotion or qua emotion-types are at times rational and at times irrational. The antecedent of this conditional expresses what I refer to as the *criterion for the ontological rationality of emotions* (COR^{e}): For *emotion* or *an emotion* there exists some normative standard that is given by what emotion or an emotion is against which our emotional responses can be judged or evaluated in virtue of the fact that our emotions manifest our rationality, i.e., the capacity for being both rational and irrational.

I argue in the remainder of this paper that the COR^{e} can be fulfilled. In doing so, I aim to explain why emotional creatures are at times rational and at times irrational in light of their emotional experiences—the rationality of emotions qua emotion or qua emotion-types. I begin by differentiating concerns regarding the rationality of emotions qua emotion from other concerns involving emotions and judgments of rationality. More specifically, I differentiate concerns regarding the rationality of emotions qua emotion or qua emotion-types from concerns regarding the instrumental rationality, the epistemic rationality, and the evaluative rationality of emotions[7]. I conclude with a discussion of how emotions qua emotion or qua emotion-types fulfill the COR^{e}, and I draw out the implications of this conclusion on the following: 1) our understanding of the possibility of human infants, some people with disabilities (who might be thought of as otherwise), and emotional animals as being rational and bearers of knowledge, 2) a way to adjudicate between internalistic, evidentialist and externalistic, reliabilist accounts of justification for knowledge, and 3) our understanding of the place of the physical in regard to what is rational and how something is known.

3. The Instrumental Rationality of Emotions

The rationality of emotions has been widely addressed by philosophers, psychologists, sociologists, and economists. We can understand many of these concerns to be about the *instrumental rationality of emotions*—about the role emotions play in rational decision-making rather than the rationality of emotions in-themselves. To illustrate this point, consider Tversky and Griffin's (1994) study of assessments of well-being based on a model of choice compared to a model of judgment. Tversky and Griffin found that assessments of well-being based on a model of choice often diverge from assessments of well-being based on a model of judgment. In their study, Tversky and Griffin gave students a scenario with two options: (A) a job paying $35,000, where equally qualified workers would receive $38,000, and (B) a job paying $33,000, where equally qualified workers would receive $30,000. Approximately half of the students surveyed were asked which job they would *choose* and the other half were asked at which job they would be *happier*. Eighty-four percent (84%) of the students who were asked

7 Note that this is not an exhaustive list of the ways in which emotions can be understood to be rational or irrational (e.g., *see* D'Arms and Jacobson (2000) for a separate although overlapping discussion about the rationality or appropriateness of emotions).

the first question chose option (A), the job that offered the higher salary compared to the other option but also gave an even higher salary to other equally qualified employees. Sixty-two percent (62%) of the students who were asked the second question expected that they would be happier with option (B), the job that offered the comparatively lower salary but also gave an even lower salary to equally qualified employees.

For our discussion here, we can understand assessments of well-being based on judgments of satisfaction or happiness as *emotion-based judgments* and we can understand assessments of well-being based on choices as *utility-based judgments*[8]. Given this, one can suggest that, in consideration of the results provided by Tversky and Griffin, at least some of these emotion-based judgments involve "irrational" emotions. One may argue that given a within-subject design, emotion-based judgments of well-being would be deemed rational if the students who chose option (A) also determined that they would be happier with option (A). The emotion-based judgments of students who would choose divergent responses or who would choose option (B) for both options would be deemed irrational since in both cases the standard for being rational would typically be set by the standard of rationality for utility-based judgments (which is typically taken as utility maximization cashed out in terms of monetary values). Therefore, emotions would be taken as irrational either because the judgments that follow from them inconsistently reflect the rational conclusions of rational utility-based judgments or because they never reflect the rational conclusions of rational utility-based judgments.

There are two points I want to make in response to this kind of argument; both of my points suggest that one may be making a category mistake when putting forth arguments like the above. First, one may be confounding two separate rational decision-making processes. Emotion-based judgments may have their own standard of rationality. So it may be problematic to contrast emotion-based judgments with utility-based judgments, and then hold rational utility-based judgments as the standard for rational emotion-based judgments. Second, even if we grant that emotion-based judgments are always irrational to the extent that they fail to consistently reflect the same outcomes given by rational utility-based judgments, this does not entail that emotions are themselves always irrational. Concerns regarding the rationality of emotion-based judgments are *not* the same as concerns regarding the rationality of emotions themselves. Concerns regarding the rationality of emotion-based judgments are concerns about the instrumental role emotions play in rational decision-making, i.e., they are concerns about the *instrumental rationality of emotions*[9].

4. The Epistemic Rationality of Emotions

Emotions can also be said to be rational or irrational in virtue of what Calhoun refers to as 'epistemic subjectivity', which presupposes a counterpart notion of epistemic objectivity (Calhoun 2004). As Calhoun notes, characterizing emotions as epistemically subjective casts them as being unreliable in terms of either failing to accurately reflect objective facts about the world or as being entirely incapable of reflecting any kind of objective fact about the world (Calhoun 2004, 108-109). This is because if we understand emotions to be epistemically subjective, and thus irrational in this sense, then emotions must meet a standard of epistemic objectivity in order to be regarded as rational. If along with cognitive theorists of emotion we grant that emotions are types of beliefs or judgments, it is generally accepted that epistemic objectivity requires emotions to be true and well warranted in order for emotions to be

8 Some might argue that this distinction may be significantly related to dual-process models of moral judgment. See Greene (2007). Also see Green (2009) for discussions regarding controversies over dual-process models of moral decision-making.

9 Also see Pollock's (2006) account of rational decision-making, especially the role of state-likings; however, one ought to note that it is not clear how state-likings are related to emotions on Pollock's account.

regarded as rational. In other words, like beliefs, the *ideal* of epistemic objectivity requires emotions to be forms of knowledge in order to be regarded as rational just as the ideal of epistemic objectivity requires beliefs to be forms of knowledge in order for beliefs to be regarded as rational[10].

Framed within what has been characterized as traditional noncognitive theories of emotion—what Deigh referred to as "feeling-centered" theories, which take emotions to be simple impressions that lack any representational content (Deigh 1994: 825)—emotions are simple feelings. In light of this, emotions would lack any kind of meaning or propositional content, and if rational assessment requires the object of assessment to be capable of conveying some meaning or propositional content that is to be assessed, then emotions would be beyond the purview of any rational assessment. They could never be rational or irrational.

These criticisms regarding the epistemic subjectivity of emotions—in virtue of relying on false or unjustified beliefs, or being incapable of representing any cognitive content—are about the *epistemic rationality of emotions*.

5. The Evaluative Rationality of Emotions

Emotions are at times rational and at times irrational in accordance with the biographical subjectivity of emotions, which refers to the fact that emotions reflect personal histories and values (Calhoun 2004: 113). The significant idea here is that emotions reflect a personal point of view. Calhoun notes that this notion of biographical subjectivity is often used to undercut the epistemic significance of emotions (Calhoun 2004: 108-109)[11]. To do so, biographical subjectivity is contrasted with a notion of biographical objectivity, and so what is rational, as arriving at truths from a biographically objective perspective[12]. From a biographically objective perspective, our view of the world is to be purified from our personal biases—from our personal histories and values.

Calhoun argues that the biographical subjectivity/objectivity of emotions may be distinguished from their epistemic rationality (epistemic objectivity/subjectivity) (Calhoun 2004: 115). Given this difference, we can understand the rationality of emotions in a sense that is distinct from concerns regarding the epistemic rationality of emotions. Judgments about the rationality of emotions when placed only against the background distinction between biographical objectivity and biographical subjectivity may be judgments about whether or not one's emotional responses from a biographically subjective viewpoint are consistent with the emotional responses from one's biographically objective viewpoint. In this sense, one would hold one's emotional response from the biographically objective viewpoint to set the standard for what counts as a rational response. In some cases, however, the standard for a rational response may be upheld by the biographically subjective viewpoint. For example, feelings of love, attachment, and jealousy, may all be rational *only* from a biographically subjective viewpoint (making such emotional responses irrational from a biographically objective viewpoint). Such cases of rationality or irrationality, which take into consideration the biographically subjective and biographically objective viewpoints as the backdrop against which our emotional responses ought to be judged as being rational or irrational, are cases regarding the *evaluative rationality of emotions.*

10 Note that what is understood to be 'knowledge', within the context of the notion of epistemic rationality, is demarcated by the traditional criteria of knowledge as justified, true belief.

11 This notion of biographical subjectivity involves what Helm would refer to as a commitment to import (Helm 2010: 310).

12 The notions of a biographically objective viewpoint and a biographically subjective viewpoint can be understood as being respectively synonymous with Parfit's notions of an 'impartial point of view' and a 'personal point of view' (Parfit 2013: 40-41, 133).

6. The Rationality of Emotions Qua Emotion or Emotion-Types as Superordinate Inference Rules

A fourth way of understanding the rationality of emotions is to understand emotions as *superordinate inference rules*. Emotions as superordinate inference rules offers a way of understanding the rationality of emotions that characterizes emotions to be arbiters of rationality rather than objects of rational assessment. Emotion-types, along with token experiences of these types, are rational *qua emotion* as a category of rational operations that implement a more general *category-level superordinate inference rule* (at the level of emotion as a *genus*). Token experiences of emotion-types are rational or irrational in-themselves—*qua the type of emotion that they are*—in virtue of the reliable operation of *type-level* superordinate inference rules (at the level of *species* of emotion).

In regard to exactly what emotion is as a category-level superordinate inference rule (e.g., emotion) or exactly what an emotion as a type-level superordinate inference rule is (e.g., as fear, anger, joy, etc.), my answer is that these are some of the questions that are currently under investigation by emotion theorists and researchers across disciplines; however, whatever emotion or an emotion is (including concerns about its status as an objective kind or a subjective kind)[13], what constitutes its essential nature as a category-level or type-level superordinate inference rule is that which is intended by speakers of some language, especially ordinary or natural languages, to be tracked by or reflected by the intentional content of their use of relevant emotion term(s), i.e., the intentionality of emotion or an emotion-type[14]. This understanding of what an emotion or emotion is as a super-ordinate inference rule is consistent with evolutionary psychological theories of emotions as superordinate programs (Cosmides and Tooby 2000), with basic emotion theories of emotions as affect programs (Ekman and Cordaro 2011), with component process theories of emotions as emergent processes (Scherer 2009), and with psychological constructionist theories of emotions as emotional meta-experiences (Russell 2008) or emotional episodes (Barrett et al. 2015)[15]. Prinz also suggests something similar with his notion of calibration files, although Prinz identifies calibration files as aspects of the initiation pathways of emotion rather than as aspects of emotion or the emotion system (Prinz 2004: 235-238).

Emotions as superordinate inference rules are thus origins of rationality. In this respect, the rationality or irrationality of token experiences of emotion can only be appropriately assessed in terms of the normative conditions that emotion or an emotion as a superordinate inference rule provides. Emotion or an emotion as a superordinate inference rule runs outside considerations of traditional or standard logical systems that dictate how assessments of warrant, rational thoughts, and rational judgments are to be evaluated. Emotion or an emotion as a superordinate inference rule is instead on par with what is typically regarded as that which functions to produce rational thought and judgment in accordance with logical rules of inference.

13 See Mun (2016) for a more detailed discussion on objective kinds and subjective kinds.

14 One might differentiate the intention that is to be tracked by or reflected by the intentional content of one's use of an emotion term from the intentionality of an emotion. My response to this distinction is that it requires further thought and research.

15 Note that emotions as superordinate inference rules may be implied by Hochschild's (2012) feeling rules, but Hochschild's feeling rules go beyond the limits of what emotions as superordinate inference rules are intended to capture. For example, corporate mandates on how their employees ought to emotionally interact with customers or clients are parasitic on emotions as superordinate inference rules rather than being superordinate inference rules of emotion. Also note that we can understand D'Arms and Jacobson's (2000) notion of the fittingness of emotions to presuppose a notion of emotions as superordinate inference rules. In order for an emotional response to be understood as fitting we must first presuppose that emotions are such things that can be fitting. D'Arms and Jacobson explain this by suggesting that our emotional responses have certain evaluative features (D'Arms and Jacobson 2000: 72). I agree, but I add here that what determines which evaluative features are the relevant evaluative features for a particular emotional response is determined by an understanding of emotions as superordinate inference rules.

Furthermore, if we consider the fact that the epistemic rationality of emotions depends on the truth of or the justification for experiencing a particular emotion, as well as what makes it possible for an emotion to be true or justified, then understanding emotions as superordinate inference rules is that which allows for the possibility of the epistemic rationality of emotions. If we understand the epistemic status of token emotions in terms of an internalistic, evidentialist account of warrant (or justification), then understanding emotions as superordinate inference rules allows one to recognize one's evidence *as warranting* (or *as justificatory*) evidence, i.e., they allow one to recognize one's evidence *as evidence* or *as reasons that warrant one's emotional response*. In this sense, superordinate inference rules can be understood as capacities that allow us to make such recognitions.

It should also be noted that this account of emotions as superordinate inference rules does not require one to posit that one must be aware of or is able to report the superordinate inference rule that allows one to recognize what counts as evidence or reasons for their emotional experience. Thus the notion of emotion or emotion-types as superordinate inference rules suggests that an internalistic, evidentialist justification of warranted emotions ultimately depends on a reliabilist meta-justification[16].

Given all of the foregoing, we can conclude that understanding emotions as superordinate inference rules allows us to understand emotions as ways of making sense of our phenomenal experiences of one's self, others, and the world. Emotions are ways of *sense-making*, and as such, provide a mode of understanding and knowing one's self, others, and the world we inhabit. I refer to this principle as *the fundamental epistemic principle of emotion* (K^e): Emotions qua emotion or emotion-types as superordinate inference rules are modes of sense-making, and as such emotions can be vehicles of knowledge.

7. Conclusion

To say that emotions qua emotion or qua emotion-types as superordinate inference rules are rational or irrational is to say that there is some normative standard that is given by what emotions are as a genus or a species (regardless of their status as an objective kind or subjective kind), against which our emotional responses are assessed or evaluated as being either rational or irrational. Thus the identification of such standards would allow us to conclude that emotions are at times rational and at times irrational—not only in terms of their instrumental rationality, epistemic rationality, and evaluative rationality, but also in-themselves, qua emotion or qua emotion-types as superordinate inference rules. Thus emotions as superordinate inference rules fulfill the COR^e. The COR^e was introduced in section §1, and states that for *emotion* or *an emotion* there exists some normative standard that is given by what emotion or an emotion is against which our emotional responses can be judged or evaluated in virtue of the fact that our emotions manifest our rationality. It is the fulfillment of this criterion regarding the ontological rationality of emotions that allows us to conclude that emotions are at times rational and at times irrational qua emotion or qua emotion-types. Emotions as superordinate inference rules fulfill this criterion.

Given that emotions as superordinate inference rules grounds the rationality of emotions qua emotion or qua emotion-types, this understanding of what emotions are challenges us to reconsider the way in which emotions are often placed in opposition to what is rational. If emotions can be intrinsically rational, and by this we mean that our emotional responses can be rational in-themselves, then there must be a conceptual space of rationality that out-runs

16 This conclusion may be taken as tracking the underlying distinction between Sosa's (2009) notions of animal and reflective knowledge, as well as presupposing a basic notion of knowledge as a state of mind. See Williamson (1995).

the conceptual space in which we place that which is rational in opposition to our emotions[17]. This conclusion allows us to understand human infants, some people with disabilities, and non-human emotional animals as being rational, although this way of being rational may be significantly different from other ways in which "normal" adult human beings may be rational[18]. This conclusion also allows for such individuals to be bearers of knowledge. From an internalist perspective on justification, human infants, some people with disabilities, and non-human emotional animals may be denied the epistemic status of knowers because their emotional responses fail to be justified by beliefs or thoughts that can be accessed or reported in accordance with internalistic, evidentialist requirements. I reject this claim, and argue from an externalistic, reliabilist perspective that, at minimum, their emotional responses can garner them the status of knowers despite their failure or inability to access or report their reasons or evidence that justify or warrant their emotional responses.
Furthermore, emotions as superordinate inference rules suggest that an internalistic, evidentialist account of emotional knowledge, wherein one is able to access or report the reasons for their emotional responses, may ultimately depend on an externalistic, reliabilist justification of emotion or emotion-types as superordinate inference rules[19]. Given this, and the assumption that emotions necessarily have physical components, it also follows that our rationality can be physically manifested and so knowledge can have a physical form.

REFERENCES

Baker, J. (2008), "Trust and Rationality", In Sosa, E. Kim, J. Fantl, J., and McGrath, M. (eds.), *Epistemology: An Anthology*, Blackwell Publishing, Malden, 1987 (first print), pp. 805-814;
Barrett, L. F., Wilson-Mendenhall, C. D., and Barsalou, L. W. (2015), "The Conceptual Act Theory: A Road Map", In Barrett L. F., and Russell, J. A. (eds.), *The Psychological Construction of Emotion*, Guilford Press, New York, pp. 83-110;
Calhoun, C. (2004), "Subjectivity and Emotions", In Solomon, R. C. (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions,* Oxford University Press, New York, pp. 107-124;
Cohen, S. (2002), "Basic Knowledge and the Problem of Easy Knowledge", *Philosophy and Phenomenological Research*, Vol. 65, No. 2, pp. 309-329;
Cosmides, L., and Tooby, J. (2000), "Chapter 7: Evolutionary Psychology and the Emotions", In Lewis, M., and Haviland-Jones, J. M. (eds.), *Handbook of Emotions, 2nd Edition*, Guilford Press, New York, pp. 91-115;
D'Arms, J., and Jacobson, D. (2000), "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophy and Phenomenological Research*, Vol. 61, No. 1, pp. 65-90;
Deigh, J. (1994), "Cognitivism in the Theory of Emotions", *Ethics*, Vol. 3, No. 4, pp. 825-854;
Ekman, P., and Cordaro, D. (2011), "What is Meant by Calling Emotions Basic", *Emotion Review*, Vol. 3, No. 4, pp. 364-370;
Fantl, J., and McGrath, M. (2008), "Evidence, Pragmatics, and Justification", In Sosa, E. Kim, J. Fantl, J., and McGrath, M. (eds.), *Epistemology: An Anthology*, Blackwell Publishing, Malden, 2002 (first print), pp. 742-759;
Greene, J. D. (2007), "Why VMPFC are More Utilitarian? A Dual-process Theory of Moral Judgment Explains", *Trends in Cognitive Sciences*, Vol. 11, No. 8, pp. 322-323;
_____ (2009), "Dual-process Morality and the Personal/Impersonal Distinction: A Reply to

17 One might argue that Baker's (2008) notion of friendship trust and the kind of rationality captured by Fantl and McGrath's (2008) cases of pragmatic justification share this conceptual space with emotions.
18 See Sosa (2009) distinction between animal and reflective knowledge.
19 One might run such an argument along the lines of Cohen's (2002) argument for animal knowledge.

Langdon, Coltheart, and Mackenzie", *Journal of Experimental Social Psychology*, 45, pp. 581-584;
Helm, B. (2010), "Emotions and Motivation in Neo-Jamesian Account", In Goldie, P. (ed.), *The Oxford Handbook of Philosophy of Emotion*, Oxford University Press, New York, pp. 303-323;
Hochschild, A. (2012), *The Managed Heart*, *3rd Edition*, University of California Press, Berkeley, 1983 (first print);
Lewis, M. & Haviland-Jones, J. M. (eds.) (2000), *Handbook of Emotions, 2nd Edition*, Guilford Press, New York;
Mun, C. (2016), "Natural Kinds, Social Constructions, and Ordinary Language: Clarifying the Crisis in Science of Emotion", *Journal of Social Ontology*, forthcoming;
Nussbaum, M. C. (2001), *Upheavals of Thought: The Intelligence of Emotions*, Cambridge University Press, New York;
Parfit, D. (2013), *On What Matters, Volume I*, Oxford University, New York;
Prinz, J. J. (2004), *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press, New York;
Pollock, J. L. (2006), *Thinking About Acting: Logical Foundations for Rational Decision Making*, Oxford University Press, New York;
Rorty, A. O. (1985), "Varieties of Rationality, Varieties of Emotion", *Social Science Information*, Vol. 24, No. 2, pp. 343-353;
Russell, J. A. (2003), "Core Affect and the Psychological Construction of Emotion", *Psychological Review*, Vol. 10, No. 1, pp. 145-172;
_____ (2008), "In Defense of a Psychological Constructionist Account of Emotion: Reply to Zachar", *Journal of Theoretical and Philosophical Psychology*, Vol. 28, No. 2, pp. 423-429;
Scherer, K. R. (2009), "Emotions Are Emergent Processes: They Require a Dynamic Computational Structure", *Philosophical Transactions: Biological Sciences*, Vol. 364, No. 1535, pp. 3459-3474;
Sosa, E. (2009), *A Virtue Epistemology*, Oxford University Press, Oxford;
Touroutoglou, A., and Barrett, L. F. (forthcoming), "How Do Emotion and Cognition Interact?", In Fox, A. S., Lapate, R. C., Shackman, A. J., and Davidson, R. J. (eds.), *The Nature of Emotion: Fundamental Questions*, *2nd Edition*, Oxford University Press, New York, pp. 1-13;
Tversky, A., and Dale, G. (1991), "Endowment and Contrast in Judgments of Well-Being", In Strack, F. Argyle, M., and Schwarz, N. (eds.), *Subjective Well-Being: An Interdisciplinary Perspective*, Pergamon Press, New York, pp. 101-118;
Williamson, T. (1995), "Is Knowing a State of Mind?" *Mind*, Vol. 104, No. 415, pp. 533-565.

PAUL POENICKE
University at Buffalo
paulpoen@buffalo.edu

EMOTIONAL INTELLIGENCE AS AN INTELLECTUAL VIRTUE: THEORETICAL ANALYSIS AND EMPIRICAL ASSESSMENT

abstract

Virtue theory has long recognized the significance of emotion for cognition, yet little philosophical research has been dedicated to identifying an intellectual virtue related to emotion. Applying recent work in virtue epistemology reveals emotional intelligence (EI) to be an intellectual virtue, as analysis demonstrates EI's ability to enhance cognition. High EI persons better attend to epistemically-significant features of the environment, which could explain the significance of stakes for knowledge attribution. While testing did not confirm higher EI with stakes sensitivity (the hypothesis), study methods, including stakes vignettes, inadvertently caused the hypothesis to be disconfirmed.

keywords

emotional intelligence, virtue epistemology, experimental philosophy, stakes, knowledge attribution

Phenomenology and Mind, n. 11 - 2016, pp. 58-67
DOI: 10.13128/Phe_Mi-20106
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

It is not a fluke that Daniel Goleman's *Emotional Intelligence* begins with an introduction entitled "Aristotle's Challenge," that itself, appropriately, opens with the philosopher's words concerning "being angry with the right person, to the right degree, at the right time, for the right purpose, and in the right way" (Goleman 1994: ix). Aristotle's words reflect virtue theory's traditional appreciation of emotion, and a recent volume advancing virtue epistemology, Robert C. Roberts and E. Jay Wood's *Intellectual Virtues*, has much to add to this legacy of emotional analysis, particularly to improve theoretical and experimental research on emotional intelligence (EI), the focus of this paper. Section I is devoted to outlining emotional intelligence, in particular John Mayer and Peter Salovey's four branch model, which defines EI as the ability to perceive, understand, utilize, and regulate emotion (Meyer and Salovey 1997: 399). After defending the Mayer-Salovey model and exploring the model through an excursus into emotional deception, Section II introduces *Intellectual Virtues* and its definition of virtue to argue that emotional intelligence is rightly considered an intellectual virtue, and how, as an intellectual virtue, emotional intelligence is related to epistemic success through emotion's epistemically-relevant features and functions.

Section III reports a study that tested the significance of emotional intelligence as an intellectual virtue. Professional philosophers and non-philosophers disagree upon the significance of stakes for knowledge attribution, and emotional intelligence, with its ability to focus the mind upon epistemically-significant features of an individual's environment, was hypothesized as a key factor in whether a participant was sensitive to changes in stakes and the resulting impact upon their attribution of knowledge. After correlating each participant's response to low and high stakes versions of popular vignettes in the literature (BRIDGE, BANK, and PINE NUTS) and that participant's emotional intelligence, results were mixed. Given the extreme length of the study and problems with the vignettes, one cannot disconfirm EI as an intellectual virtue based upon the empirical data. Emotional intelligence emerges from this paper as a novel virtue with a promising foundation based on a combination of theory and experiment.

1. A Brief Overview of Emotional Intelligence

While the term 'emotional intelligence' has been used since the mid-20th century, psychological research concerning the concept did not appear until the 1990s. Inspired by the work of John Mayer, Peter Salovey, and others, Daniel Goleman's *Emotional Intelligence* was published in 1995 and became a best seller, prompting a flood of popular works and encouraging further academic research. Within the psychological community, the nascent field of emotional intelligence has split between ability and skill based models. Ability based models are devoted solely to measuring mental abilities, such as recognition and knowledge, regarding emotion and emotion's interaction with cognition; mixed models take a broader view of emotional intelligence, consider emotional intelligence as a single entity that has intellectual, social, and motivational aspects (Schulze and Roberts 2005: 399-405).

Which model of emotional intelligence is best? An ability model, particularly the four branch Mayer and Salovey model, has several distinct advantages[1]. Ability models focus on the 'intelligence' of emotional intelligence, directing their attention to those emotion-related aspects of cognition omitted by traditional views of intelligence. Mixed models tend to see emotional intelligence as a wide range of skills and abilities that extend far beyond ordinary and technical definitions of intelligence, muddying the meaning of 'intelligence.' By focusing on the interaction between emotion and cognition—emotional perception, regulation, and reasoning—research is far more fruitful: comparisons with traditional intelligence can be directly ascertained, complicated questions regarding emotions and their impact on behaviors and skills can be avoided, and fruitful tests measuring emotional intelligence can be constructed. While the focused study of emotional intelligence is only in its second decade, evidence leads many to believe that Mayer and Salovey's model, and ability models generally, has the greatest evidential support among models of emotional intelligence, which has made it the favorite of both supporters and critics of emotional intelligence (Schulze and Roberts 2005: 31-51, 309-342; Matthews, G., Zeidner, M., and Roberts 2002: 31-79, 511-551). Therefore, it seems that if we want to utilize a technical definition of emotional intelligence, then the work of Mayer and Salovey should be given primacy.

Mayer and Salovey's understanding of emotional intelligence is broken down into four branches or levels of skills or abilities. First, there is the *perception* and *expression* of emotion, as one identifies and expresses emotions subjectively and in extra-personal contexts, whether in others, a piece of art, or an animal[2]. Second, emotion is *assimilated* into thought. A person is able to weigh diverging emotions, consider when emotions seem inappropriate in a context, and directly attend to and compare emotions—say, examining one's subjective fear or excitement regarding two different events. *Understanding* and *reasoning* about emotions, the third skill, allows us to cognitively interact with emotions by labeling emotions and recognizing the difference between phenomenologically similar feelings (such as anticipation and nervousness), appreciating how emotions can change (from extreme fear to relief after the resolution of a frightening episode), and grasping the connection between our emotions and various contexts. The final level of emotional intelligence, its apogee, involves the reflective *regulation* and *management* of emotions. This includes being able to monitor one's own emotions and promote emotion and intellectual growth, as well as helping to regulate another's emotions (Mayer and Salovey 1997: 396-417).

A brief excursus concerning emotional deception will help to further explain the model and

1 Mayer, Salovey, and Caruso (2000), summarizes the previous work done by Mayer and Salovey.

2 Roberts and Woods would emphasize that perception is a form of acquaintance, where one is either in immediate cognitive contact with something or has had such contact and still retains that information through memory, understanding, or another ability (Roberts and Wood 2007: 50-51).

highlight the importance of emotional intelligence. Emotions routinely deceive: irrational fears, especially those that persist even after we grasp reasons for not being afraid, exist as unpleasant reminders that emotions routinely trick us into holding false beliefs. However, if we have emotional intelligence, then we can anticipate the myriad of ways emotions can lead us astray and, through various methods, discourage its disruption of rationality. Such emotional subterfuge is not necessarily difficult to detect, if time and effort is made to understand our own psychologies. Gaining an initial perception of our emotions and their dispositions to distort might come from meditation, where individual emotions could be felt and analyzed; reading literature, where we could see how others' emotions lead to harmful effects and appreciate how emotionally mature characters recognized their emotions; or journaling, where one writes out what one is feeling, considers the emotion in question, and then reassesses one's emotional experiences. A person could also read books like *Emotional Intelligence* to further enhance their understanding of emotions; they could take tests, such as Mayer and Salovey's Multifactor Emotional Intelligence Scale Study (MSCEIT), to see how well he can recognize and analyze emotions. Once emotions are perceived, we can then begin to express them, which will promote further perception of emotion in ourselves and others, allowing us to recognize and overcome emotions' occasional invisibility. This initial step would allow us to appreciate emotional deception, but it would be further work that would be refined as emotional analysis and understanding (the third level of Mayer and Salovey's model), that would substantially reduce and correct emotional deception. Such refining would be especially critical to remove emotional triggers, and include an appreciation of how emotions can change, an investigation into the causes and results of certain emotional reactions, an exploration of the differences and similarities in emotions, and a realization of how moods, feelings, and emotions are connected. With such emotional analysis and understanding in place, a person can finally focus on how to regulate emotions—what management skills are necessary to reduce the impact of harmful emotions, how to encourage helpful feelings, and how to stay open to emotional experiences in ways that encourage intellectual and emotional growth (Mayer, Salovey, and Caruso 2000a: 400-401).

2. Emotional Intelligence: A Virtue

In *Intellectual Virtues*, Roberts and Wood propose that virtue is, "...an acquired based of excellent functioning in some generically human sphere of activity that is challenging and important" (Roberts and Wood 2007: 59). Salovey and Meyer's definition of emotional intelligence "...as the ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion in thought, and regulate emotion in the self and others" (Meyer and Salovey 1997: 399-401) is an intellectual virtue in the eyes of Roberts and Wood. An initial doubt might derive from Salovey and Meyer's assertion that emotional intelligence is an ability, not excellent functioning, although, given Salovey and Mayer's philosophically non-technical, fairly loose use of 'ability,' this definition of emotional intelligence as an ability could probably be expanded to include excellent functioning as well. Additionally, Roberts and Wood assert that many virtues are dispositions of the will, including emotions; further, as shown in the above discussion regarding emotional deception, emotional intelligence would be a virtue, since it is a disposition that "...correct(s) for proneness to dysfunction and error in certain common situations" (Roberts and Wood 2007: 59). One might also argue that emotional intelligence is a particular faculty—the innate or natural "powers" possessed by every human being—whose maturation, development, and perfection has made it into a state of virtue (Roberts and Wood 2007: 59). For a variety of reasons, emotional intelligence would appear to count as a virtue for Roberts and Wood. With this in mind, the rest of the definition is easily achieved: emotional intelligence is certainly an acquired, not inherited or innate, base of functioning; it is not used only in a specialized, remote context but rather on

a wide variety of occasions; and it is certainly challenging to acquire and critical to possess. Emotional intelligence fulfills other characteristics Roberts and Wood find in virtues—virtues are correctives for human defections, the product of a formation in human excellence or "education for life" (Roberts and Wood 2007: 69, 165). Thus, it seems as if Roberts and Wood would include emotional intelligence as a "disposition of intellectual excellence" that is necessary to correct the "...perennial set of deficiencies which in every generation need to be corrected" (Roberts and Wood 2007: 22).

It is clear that Roberts and Wood appreciate the importance of emotions, and *Intellectual Virtues*, in chapters on practices and practical wisdom, further explores the value of emotional intelligence. Roberts and Wood emphasize that good reading requires an appreciation and appropriation of emotions associated with Derridaian and Gadmerian reading styles—a sense of creativity and joy balanced with a concern for respect and diligence (Roberts and Wood 2007: 123-133). Public debate requires generosity to appreciate another's position, meekness to encourage a free and welcoming discussion, and friendliness to enjoy protracted disagreements over fundamental issues (Roberts and Wood 2007: 134-140). Teaching requires patience with students' sometimes offbeat questions, harshness to overcome intellectual sloth, and honesty to represent all positions fairly; a student needs to be open to learn new things and enthusiastic to discover new truths. Intellectual practices require emotional regulation and promotion, critical aspects of emotional intelligence.

These facts would be included in practical wisdom, Aristotle's *phronesis*, the intellectual side of virtue that focuses reason and judgment on these matters of practice. Practical wisdom is both deliberative, the active planning of a solution, and perceptive, passively taking in a context and its relevant facts. Perceiving, assimilating, understanding, and regulating emotions—the four skills of the Mayer-Salovey model—are critical to both the deliberative and perceptive powers of practical wisdom, as Roberts and Wood's example of the college president demonstrates[3]. During apartheid, a college president wants to show solidarity with South African blacks by hosting an open debate concerning the appropriateness of divesting university funds from South Africa as a manner of promoting justice. Practical wisdom allows him to perceive that his students might be angered by such debate; that a difficult that debate is critical for intellectual honesty; and that his own proclivity for intellectual cowardice has kept him from acting in accord with his love of knowledge in the past. Practical wisdom also allows the president to take active control of his emotions through self-management strategies, to empathize with certain portions of the college community, and how to make an open debate an inviting, non-threatening place to voice opinions. Without emotional intelligence, practical wisdom may be blind to helping the president appropriately perceive and act in this situation.

Beyond its impact upon practical rationality, emotional intelligence indirectly impacts belief and knowledge through its regulation and management of emotions, which, according to Georg Brun and Dominique Kuenzle, are highlighted by five epistemically-relevant features and functions (Brun and Kuenzle 2008: 16-22). First, emotions provide enhanced motivational force for epistemological study: surprise, doubt, and puzzlement promote further reflection on epistemological questions, encourage rigorous research, and force us to focus on troubling issues and arguments. Without emotions such as doubt, many philosophical projects, from Socratic ignorance to Cartesian skepticism, may never have been initiated. Second, emotions help us determine salience and relevance. Philosophy takes place within a sphere of limitation: the beings that conduct philosophical investigation cannot pursue every line of inquiry,

3 Roberts and Woods: 138-140, 313-315.

question, or objection; emotions are spotlights that, encompassing our attitudes, concerns, and beliefs, focus debate on what is significant and pertinent. Third, emotions offer access to facts and beliefs, either in a unique—that is, it is solely because of emotions that we have such epistemic access—or dispensable manner, where other states could provide such access. Emotions could also contribute in non-propositional ways to knowledge and understanding. This fourth feature would include the arrangement of categories and other forms of cognitive organization, as well as standards of inquiry and warrant. Finally, emotion allows for epistemic efficiency: emotion certainly makes it easier to perform certain epistemic tasks, as intuitive thinking that relies upon emotion allows for quick, accurate responses to various environments.

3. Testing Emotional Intelligence as an Intellectual Virtue

Simply identifying emotional intelligence as a virtue and considering its general significance for epistemology falls short of assessing whether EI actually functions as a virtue to resolve a genuine epistemic controversy. This section discusses a study that tested the significance of emotional intelligence as a key cause of sensitivity to stakes effects on knowledge attribution. Until the late 1980s, philosophers investigating the conditions for knowledge attribution presumed that such conditions are based upon truth-relevant factors, those elements which are able to increase or lower the likelihood of the subject's belief that p is true – e.g. justification, reliability, and evidence. In the last few decades, several researchers started to inquire whether features of the conversational context and truth-irrelevant factors, such as the subject's error possibilities and other practical interests, might affect knowledge attributions as well. Contextualists (DeRose 2009) argued for a controversial thesis: in certain conversational contexts, particularly in those situations where error possibilities have been raised, it may no longer be true that an individual knows, even if that individual knew in contexts when those error possibilities had not been raised. Responding to the contextualist challenge included Jason Stanley (Sripada and Stanley 2012), who argued that practical interests matter as an epistemic fact about a situation. According to Stanley's interest-relative invariantism (IRI), whether we should attribute knowledge in a given situation is not based on the conversation context, as suggested by contextualists; rather, "the truth conditions of a sentence like 'S knows that p' fluctuate according to what is at stake for the subject" (Buckwalter 2010: 396).

Previous experiments have revealed a surprising insensitivity on the part of ordinary subjects to stakes in various experiments attempting to confirm contextualism or IRI[4]. If emotional intelligence is a virtue and hones the epistemic advantages of emotions, then EI could provide the kind of epistemically-significant sensitivity towards stakes that has yet to be discovered. Emotions aid individuals in determining the salience and relevance of reasons and evidence. Emotion is uniquely connected with action, as emotions provide motivation through both fast responses to immediate stimuli and slow responses to quotidian issues via engrained moods, propensities, and regulation. The raising and lowering of stakes would be of great significance for practical rationality, which, as discussed in Section II, requires emotional intelligence. This sensitivity to changes in stakes would appear in testing through a number of different features of the high EI individual. The emotionally-intelligent person would have an enhanced ability to use emotional perception, regulation, and reasoning to better motivate themselves to analyze and answer confusing vignettes, providing researchers with determined, engaged participants. The emotionally intelligent person is also more empathic, which will make it easier for that individual to insert themselves into a vignette and determine the impact of

4 See Buckwalter and Schaffer (2013) for a history of the experimental literature on stakes.

stakes on that individual's epistemic situation. Finally, emotions allow for more ready access to intuitions and intuitive responses that are being tested in experimental philosophy.
The study aimed to uncover what kind of individual differences exist in responses to widely discussed epistemological cases between participants who score highly on tests of emotional intelligence versus participants with low scores on such tests. The main hypothesis was that high EI subjects demonstrate a higher sensitivity to the impact of stakes on evidential standards than low EI subjects.
Sixty undergraduate students at a public university participated in the study in exchange for extra credit in one of their courses. Students were predominantly Anglo-American, with roughly equal proportions of male and female students and an average age of twenty. Each student spent between forty-five to sixty minutes completing the test materials, which were given in different sequences to eliminate ordering effects. The study materials consisted of two elements, the first being Mayer and Salovey's Multifactor Emotional Intelligence Scale Study (MSCEIT), the one hundred forty-one item test composed by Mayer and Salovey to measure emotional intelligence.
The second part of the study materials featured three sets of vignettes: BRIDGE, BANK, and PINE NUTS. Each participant received three vignettes, one from each set, in which the protagonist is in either in high or low stakes. The simplest set of vignettes, BRIDGE, came from Feltz and Zarpentine (2010). The low stakes vignette is as follows:

> John is driving a truck along a dirt road toward a remote mountain village so that he can study the culture of the people that live there. He comes across a rickety wooden bridge over a shallow ditch. John thinks to himself, "The bridge should be stable enough to support the weight of my truck." He then drives the truck onto the bridge and makes it safely over to the other side.

Stakes were raised in another variant of BRIDGE by increasing the height of the bridge (in high stakes, the bridge spanned a thousand foot gorge), and then amplified in a third version by increasing the bridge's height and changing John's goal (John hauls essential medicine over a gorge to aid a remote village, where desperately ill orphans suffer from life-threatening diseases). After each vignette, with 1 being 'Strongly Disagree' and 7 being 'Strongly Agree', participants were asked: "Please indicate the extent to which you agree or disagree with the following claim: 'John knew that the rickety wooden bridge would be stable enough to support his truck'".
Participants also received a variant of BANK from Buckwalter (2010). Here is the low stakes version:

> Sylvie and Bruno are driving home from work on a Friday afternoon. They plan to stop at the bank to deposit their paychecks, but as they drive past the bank they notice that the lines inside are very long. Although they generally like to deposit their paychecks as soon as possible, it is not especially important in this case that they be deposited right away. Bruno tells Sylvie, "I was just here last week and I know that the bank will be open on Saturday." Instead, Bruno suggests that they drive straight home and return to deposit their paychecks on Saturday. When they return to the bank on Saturday, it is open for business.

Stakes were first raised in a second version of BANK by adding skeptical doubts (Sylvie suggests that banks typically are closed on the weekend, and that this bank had been closed on Saturday in the past). The final version raised stakes through adding Sylvie's skeptical

doubts as well as including the possibility of financial peril: if Bruno cannot deposit his check on time, he will be in a precarious position with his creditors. After each vignette, with 1 being 'Strongly Disagree' and 7 being 'Strongly Agree', participants were asked: "Please indicate to what extent you agree or disagree that Bruno's assertion, 'I know the bank will be open on Saturday,' is true".

Finally, the study presented a variant of PINE NUTS from Sripada and Stanley (2012). Participants received the following as the low stakes version of PINE NUTS:

> Hannah likes the taste of most foods and is not a very picky eater. One evening, Hannah and her sister Sarah are at a brand new restaurant that has just opened up. Hannah orders a plate of noodles. When her food is brought to the table, Hannah notices something that looks like nuts sprinkled on her noodles and wonders what it is. Sarah says, "The noodles may be topped with pine nuts." Hannah notes that the menu says her dish does not contain pine nuts. Based on this, Hannah forms the belief that her noodles are not topped with pine nuts.

The two additional adaptations of the vignette raised stakes by increasing Hannah's allergic reaction to eating pine nuts (from dry mouth to a deadly reaction from eating a single nut). After each vignette, with 1 being 'Strongly Disagree' and 7 being 'Strongly Agree', participants were asked two questions, to directly assay the connection between evidence and knowledge predication:

1) "What is the strength of Hannah's evidence that her noodles are not topped with pine nuts?"
2) "Suppose it turns out that her noodles are not topped with pine nuts. Please rate how strongly you agree or disagree with the following sentence: 'Hannah knows her noodles are not topped with pine nuts.'"

For each contrasting pair of thought experiments, participant responses to these thought experiments were correlated with that participant's emotional intelligence scores. Participants with certain ranges of scores were grouped together and statistical tests performed to determine if the mean responses given by each group differed significantly from one another. Effect sizes were measured and compared.

From the results of our study, we were unable to confirm or deny our hypothesis, that a participant's emotional intelligence determined whether the subject was sensitive to the impact of stakes on evidential standards for knowledge predication. Certain groups of EI aptitude scored on the stakes vignettes as predicted by our hypothesis, but this pattern confirming our hypothesis's prediction did not occur across the ranges of high and low EI and resulting changes in stakes. Our results did not allow us to reject the null hypothesis, the hypothesis that there is no relationship between EI and sensitivity to the impact of stakes on evidential standards for knowledge. Furthermore, we were unable to establish a link between EI and sensitivity to stakes based on the significance of goals, the addition of skeptical hypotheses, and the connection between evidence and knowledge attribution. Unfortunately, this study cannot provide confirmation or disconfirmation to the contextualist or interest-relative invariantist intuitions that motivated the studies by Feltz and Zarpentine, Buckwalter, and Sripada and Stanley.

4. Lessons for the Future

What accounts for the uneven results? Two explanations appear plausible and should be considered in future work. First, mental fatigue impacted participants, who—in one sitting—completed the lengthy MSCEIT and worked through a number of torturous vignettes. A person of high emotional intelligence, an individual who could manage their emotions to promote dedication and interest, would find themselves exhausted, providing mistaken answers, after such a lengthy battery of tests. If researchers want to use the MSCEIT in future work, they should allow participants to finish the test in one session and return for additional testing on a different day.
The second explanation is that the construction of the vignettes confounded the study's attempt to compare EI and sensitivity to stakes. Buckwalter's experiments on bank cases—as well as all the studies that follow Buckwalter—repeat, without alteration, DeRose's thought experiments from the late 1990s, a decade before online banking and ATMs. The standard study participant, an undergraduate student, lives with these modern banking convinces, allowing money to be deposited unencumbered by long bank lines. Students are likely bamboozled as to why an individual would worry about bank lines when money can be deposited via an ATM, and as a result, they may be responding based on intuitions completely unrelated to the situation, stakes, and standards of evidence.
The vignettes in the stakes literature are written to test specific hypotheses but are not composed with an eye towards presenting a realistic portrait of how humans actually interact and live—the focus of an empathetic, emotionally intelligent person. In the high stakes bank case, the vignette strains credulity: Bruno's claim that he had visited the bank the previous week could be rebutted by Sylvie rattling off skeptical hypotheses—banks can change their hours and typically close on weekends—but without her responding specifically to Bruno (i.e. providing him with an estimate when in the past the bank had different weekend hours), it seems like Bruno and Sylvie are talking past one another. An emotionally-intelligent individual, using their ability to empathetically place themselves in the situation to evaluate Bruno's assertion, would find it difficult to make sense of whether Sylvie has actually provided evidence that increases the stakes.
Sripada and Stanley state that they are concerned that providing such detail would artificially focus participants' attention on stakes (Sripada and Stanley 2012: 6-9). While simpler vignettes could be constructed to nullify this concern (see BRIDGE as an example), Sripada and Stanley also worry that given high stakes, participants may attribute knowledge simply because higher stakes tend to encourage additional evidence-gathering practices, outside those mentioned in the vignette, by individuals open to severe risks (Sripada and Stanley 2012: 9-11). It is for this reason that the authors crafted the "ignorant" version of PINE NUTS utilized in this study. However, the "ignorant" high stakes situation, in which Hannah does not have a belief about her severe allergy, is dubious. In a world where pine nuts can kill, are used by restaurants and listed in menus, and can be readily perceived by non-experts, we expect Hannah to be in a position, regardless of her additional attempts to gain evidence, to have a belief about her allergy. A world with all the features mentioned in the previous sentence, and where restaurants did not protect their customers from deadly allergies caused by genetics, out of fear from the law and social sanction, is a very different world from our own. Most importantly, such a counterfactual world would strain the high EI participant from being able to empathize with the epistemic situation of the protagonist.
This concluding section highlighted the ways in which experimental artifacts—the vignettes used to assay the relationship between emotional intelligence and stakes—distorted the results of our study. Beginning with psychology, a feature of the psyche was identified that has been described by Roberts and Wood as a virtue. Emotional intelligence, while an original addition to the field of epistemology, has the benefit of being empirically grounded and tested based on philosophically-significant criteria. If research into the virtues is to bear fruit and reply to the

critiques of John Doris (who challenges the existence of such character traits in [Doris 2002]), then virtue theorists need to become familiar with empirical research, including experimental design. Simply knowing the latest science on character is not enough—philosophers need to be creative, looking to provide analysis based upon a variety of evidence and evidence-gathering techniques. This paper stands as a signpost pointing towards the fecundity of emotional intelligence and the significance of experimental philosophy for those who tread in Aristotle's footsteps.

REFERENCES

Brun, G., and Kuenzle, D. (2008), "A New Role for Emotions in Epistemology?", In Brun, G., Doguogu, U., and Kuenzle, D. (eds.), *Epistemology and Emotions,* Ashgate Epistemology and Mind Series, Ashgate Publishing Company, Burlington, Vermont, pp. 1-31;
Buckwalter, W. (2010), 'Knowledge Isn't Closed on Saturday: A Study in Ordinary Language', *Review of Philosophy a Psychology*, Vol. 1, No. 3, pp. 395-406;
Buckwalter, W. and Schaffer, J. (2013), "Knowledge, Stakes, and Mistakes", *Noûs*, Vol. 47, No. 1, pp. 1-34;
DeRose, K. (2009), *The Case for Contextualism: Knowledge, Skepticism, and Context: Vol. 1*, Clarendon Press, Oxford, pp. 304;
Doris, J. (2002), *Lack of Character: Personality and Moral Behavior*, Cambridge University Press, Cambridge;
Feltz, A. and Zarpentine, C. (2010), "Do You Know More When It Matters Less?", *Philosophical Psychology*, Vol. 23, No. 5, pp. 683–706;
Goleman, D. (1994), *Emotional Intelligence,* Paperback, Bantam Books, New York;
Matthews, G., Zeidner, M., and Roberts, R. (2002), *Emotional Intelligence—Science and Myth,* A Bradford Book, The MIT Press, Cambridge, MA;
Mayer, J., Salovey, P., and Caruso, D. (2000), "Emotional Intelligence", In Sternberg, R. (ed.), *Handbook of Intellegences,* Cambridge University Press, Cambridge, pp. 396-422;
Mayer, J. and Salovey, P. (1997), "What is Emotional Intelligence?", In Salovey, P., and Sluyter. D. (eds.), *Emotional Development and Emotional Intelligence: Educational Implications,* Basic Books, New York, pp. 3-31;
Roberts, R. and Wood, J. (2007), *Intellectual Virtues*, Clarendon Press, Oxford;
Schulze, R. and Roberts, R. (2005), *Emotional Intelligence—An International Handbook,* Hogrefe and Huber, Cambridge, MA;
Sripada, C. S. and Stanley, J. (2012), "Empirical Tests of Interest-Relative Invariantism", *Episteme*, Vol. 9, No. 1, pp. 3-26.

ELISA MAGRÌ
University College Dublin
elisa.magri@ucd.ie

AFFECTIVITY AND SELF-DISPLACEMENT IN STEIN'S EARLY PHENOMENOLOGY
ON THE ROLE OF SELF-EXPERIENCE IN EMPATHY[1]

abstract

In this paper, I shall focus on the role of bodily self-displacement in Stein's account of empathy, pointing out its relevance in the general dimension of affectivity. In my view, Stein grounds empathy on a dynamic model of embodied self-experience, which shares significant similarities with Varela & Depraz's neurophenomenology. However, I shall argue that Stein's view of empathy cannot be reduced to a naturalised phenomenological sense and that bodily self-displacement is pre-condition of a more complex disposition towards others as in line with Ratcliffe's theory of radical empathy.

keywords

phenomenology, empathy, affectivity, self-displacement, Edith Stein

Phenomenology and Mind, n. 11 - 2016, pp. 68-77
DOI: 10.13128/Phe_Mi-20107
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

From a phenomenological point of view, empathy allows a basic grasp of another's experience without producing any identification or simulation. Both in *Ideas II,* in the V *Cartesian Meditation*, and in the voluminous manuscripts on intersubjectivity (Husserliana 13-15), Husserl argues that empathy is what allows the other to be given to me as perceptually present (Zahavi 2014: 127). For Husserl, empathy primarily involves apperceiving the other as an embodied self, whose experience is passively grasped drawing on the reservoir of meaning that is provided by the experience of one's own body. In this sense, sensing one's own body plays a crucial role in that it represents the primary condition for being able to respond to another's state. As Zahavi has pointed out, «my body is given to me as an interiority, as a volitional structure, and as a dimension of sensing, but it is also given as a visually and tactually appearing exteriority. And the latter experience, the fact that my own self-experience is characterized by this remarkable interplay between *ipseity* and *alterity*, is, according to Husserl, precisely one of the elements that must be in place if empathy is to be possible» (Zahavi 2014: 135). With this regard, Zahavi stresses that bodily self-experience counts as precondition of empathy and not as a model for other-experience. When I encounter someone whose experience exceeds or contrasts with my own, my experience does not constitute any model, but rather it represents the dimension against which the other can be sensed by contrast. Thus, a more complex explication of our receptivity to the other's experience is necessary for empathy to be fulfilled.

Following Husserl, Stein developed her dissertation on the problem of empathy by taking into account the constitution of the lived body. For Stein, empathy involves both apperceiving the other as an embodied self as well as understanding the other as a person. Although these two experiences are distinct from each other, they are not separate. While the experience of one's own body is based on dynamic aspects of sensitivity and affectivity that intersect one's own subjective experience with that of another, understanding others as persons involves the appraisal of *die geistige Person,* i.e. the constitution of the person as a subject belonging to intersubjective spiritual nexuses. In the following, I will restrict my analysis to Stein's account of the lived body, which includes an original appraisal of affectivity and self-experience in the context of empathy.

With this regard, Stein's account of empathy has recently been the object of a number of

1 I wish to thank both editors and two anonymous referees for helpful comments and corrections on a previous draft.

articles and in-depth analyses[2]. Among these, Vendrell Ferran's study is remarkable in that it brings to the fore the importance of affective intentionality in Stein's early works. As Vendrell Ferran has shown, Stein proposes an original stratification of emotional life that is based on the reach, duration, and intensity of feelings[3]. In relation to this, my analysis will target an aspect of Stein's account of affectivity that is central for empathy, i.e. the role of bodily self-displacement. More precisely, I shall concentrate on the fact that one's own self-experience is not the experience of a monolithic and static subject, but rather a dynamic phenomenon, which is characterised by interrelated episodes of sensations, emotions, and bodily changes. In this respect, Stein shows that empathy presupposes the kaleidoscopic shifting of bodily self-perception, thereby anticipating several important findings of genetic phenomenology in an original way.

1. Bodily Self-Displacement

To begin with, Stein's analysis of the constitution of the lived body centres on the distinction between proprioception and heteroception. According to Stein, the body constituted in outer perception is a physical object that is out "there". Yet, unlike other objects located in space, my physical body «is always given to me in successive appearances only variable within narrow limits. As long as I have my eyes open at all, it is continually there with a steadfast obtrusiveness [*Aufdringlichkeit*], always having the same tangible nearness as no other object has. It is always "here" while other objects are always "there"» (Stein 2010a: 57; Stein 1964: 39). The reference to the "obtrusiveness" of one's own body suggests that the lived body participates in the horizon of my perceptual field, providing orientation in space. If I pay attention, I can discern a variety of bodily features that limit my visual field: for instance, the curve of my nose, the movements of my body, my posture etc. In this sense, perceiving outer objects does not coincide with perceiving oneself as having a certain position in space. Such difference corresponds to the basic distinction between proprioception, e.g. sensing one's own body as spatially oriented, and heteroception, e.g. seeing oneself in the mirror. For this reason, sensations correspond to sensory fields that are localised in the body. In contrast with judging, willing, and perceiving, sensations are «always spatially localized "somewhere" at a distance from the "I" (perhaps very near to it but never in it), I can never find the "I" in it by reflection. And this "somewhere" is not an empty point in space, but something filling up» (Stein 2010a: 57; En. tr.: Stein 1964: 39).

By referring to the spatial location of sensations and to their distance from the "I", Stein does not defend any alleged Cartesian dualism. On Stein's account, sensations are not sensory data devoid of meaning, but rather they stand for the originary sensitive motility of the body as they "fill up" the organs and generate different possibilities for action and movement. In this sense, far from introducing any hiatus between understanding and sensibility, Stein's inquiry focuses on the existence of sensory fields that are permeated with possibilities for action. From this point of view, the analysis of sensations makes explicit a form of motor intentionality that is based on the shifting of the surrounding world due to body movements.

2 For an overview of current debates on empathy that include the appraisal of Stein's phenomenology see Meneses & Larkin (2012), Aaltola (2013), Haney (2013), Dullstein (2013), Zahavi (2014: 123-145), Vendrell Ferran (2015), Jardine (2015), Lebech (2015) and Svenaeus (2015). While Meneses & Larkin, Haney, Zahavi, Aaltola and Dullstein have extensively developed the problem of perception in Stein's account of empathy, showing the significance of Stein's phenomenology for contemporary psychology, Vendrell Ferran, Jardine and Svenaeus have pointed out the relevance of affectivity in Stein's view of empathy. For a detailed reconstruction of the importance of Stein's phenomenology in contemporary debates on empathy and social ontology see also Szanto & Moran (2015).

3 Cf. Vendrell Ferran (2015), who convincingly argues that Stein's approach shares Brentano's paradigm on affectivity, but she also made important changes to it.

It follows that one has different sources of movements depending on the type of action that is at stake: «This zero point is not to be geometrically localized at one point in my physical body; nor is it the same for all data. It is localized in the head for visual data and in mid-body for tactile data» (Stein 2010 a: 58; Stein 1964: 40). If I sit and stay still on my chair, my head and my eyes will determine my sense of orientation, but in dancing and moving the body has the torso as primary source. As a result, there are as many zero points of orientation as possibilities for action and movement.

Moreover, Stein emphasises that the relation that I can establish between my own zero point and the different parts of my body cannot be compared to the distance of external objects to me. In this sense, body space (*Leibraum*) and outer space reflect a distinction between the first-person and the third-person point of view. Body space presents itself as the *locus* of motor intentionality and it is available to self-experience, while having a representation of one's own body as an item located in outer space corresponds to the third-person perspective. What Stein emphasises is the intermodal translation of different ways of sensing that is missing in the third-person perspective. I do not simply see things as items located at measurable distance from my own position, for I also *sense* them: «The robes in Van Dyck's paintings are not only as shiny as silk but also as smooth and as soft as silk» (Stein 2010a: 61; 1964: 41-42). With regard to this, Stein mentions that there are different forms of associations, which are not exclusively based on representations, for example when I see a lump of sugar and I remember that is sweet. Stein's argument is that I can *see* the sweetness of the sugar lump without entertaining a picture of it. This because my own perception of the sugar lump is embedded with emotional and synesthetic features that draw on my past experiences and influence my overall disposition to objects and events. In my view, Stein's extensive observations on the constitution of the lived body indicate that kinetic and synesthetic features are conjoined in perception in order to provide a dynamic and embodied model of self-experience.

In relation to this, it is noteworthy that Stein was familiar with the studies of Moritz Geiger, who distinguished between psychic state (*Zustand*) and psychic activity (*Tätigkeit*). While states are usually related to feelings provoked by certain objects (e.g. a melody or a painting), activities are bodily movements that can be influenced by feelings (Geiger 1911: 50). Accordingly, Geiger differentiated between empathy of moods and empathy of activities or movements. While the former is induced by an external object and rests on psychological associations, empathy of activities depends on the phenomenon of *Gefühlverschmelzung*, i.e. the fusion of feelings caused by seeing someone doing something. As pointed out by Pinotti (2000: 465), Geiger conceived of empathy as a complex phenomenon brought about by the distinction between two poles: the subject of the experience and the object, which latter induces certain states or feelings in the subject. Stein's reference to Van Dyck's painting seems a direct reference to Geiger's theory of empathy. Unlike Geiger, Stein insists that a system of intentional nexuses takes place between the body and the environment so that what is given in outer perception is not represented pictorially, but rather reactivated from the nexus of retained experiences and integrated in a general sense of motor intentionality.

Importantly, Stein's discussion of the lived body brings to light the relevance of bodily self-displacement in the psychophysical constitution of the ego. As Sheets-Johnstone (2015) has recently argued, the proprioceptive sense that I am when sitting or standing does not immediately translate in a "pre-reflective sense of myself as embodied". Quite to the contrary, the kinaesthetic experience of movement is of a particular qualitative dynamic that cannot be reduced to ownership. According to Sheets-Johnstone, ownership is «an adult concept and a third-person term, a strictly objective, even culturally derived and culturally-anchored observation [...]. Apart from the fact that what I have I may own – a car, a computer, a plot

of land – to speak of ownership of one's own movements is phenomenologically vacuous» (Sheets-Johnstone 2015: 35). In other words, Sheets-Johnstone remarks that kinaesthetic movements have the function of making us aware of the potential self-displacement of the ego rather than giving us a sense of ownership. Stein's analysis of the lived body points in the same direction: if the psychophysical constitution of the body has to provide the basis for empathy, it must also comprise a fundamental character of self-displacement and self-distantiation. For Stein, what is disclosed by empathy is not a monolithic subject without quality, but rather the fundamental quality of sensitivity as disposition to feel and to be affected in the relation to the world. In this sense, bodily feelings cannot be separated from sensations:

> Every mental act, every joy, every pain, every activity of thought, together with every bodily action, every movement I make, is sluggish and colorless when "I" feel sluggish. [...] Not only do I see my hand's movement and feel its sluggishness at the same time, but I also see the sluggish movement and the hand's sluggishness. We always experience general feelings as coming from the living body with an accelerating or hindering influence on the course of experience (Stein 2010 a: 65; Stein 1964: 45).

Somatic feelings, such as sluggishness, influence my overall sense of inhabiting a certain environment. However, only somatic feelings have the quality of filling up the body. Feelings that are not strictly somatic in nature, such as joy or melancholy do not. Still, they influence and interact with each other, and cohere together in generating my sense of being in time. Thus, feelings shape the world-view in which I am inserted. From this point of view, the stratification of emotional life addressed by Vendrell Ferran in relation to Stein's phenomenology can be further differentiated in light of Varela & Depraz's analysis of time and affectivity (Varela & Depraz 2005).

2. When Varela Meets Stein

To be sure, Stein distinguishes between different forms of feelings on the basis of their duration (Vendrell Ferran 2015). However, such distinction is a pre-condition of empathy in that it shows that affects are always embedded with a specific form of response to the other, whether the environment, an object or another subject, thereby generating a specific sense of inhabiting the world. On Stein's (and Varela's) account, there are as many affective levels as temporal phases: states represent the awareness of the tonal shift that is constitutive of the living present (they fill up the body); higher-order feelings stand for the dispositional trend of coherent sequences of embodied activities (they do not fill up the body, but give the general tone of my being in time); moods (*Stimmungen*) exist at the scale of narrative description over time. Although states, feelings, and moods all develop over time, they are qualitatively different with regard to their affective tonality as well as to their axiological properties. In this sense, in Stein's view, the givenness of one's own body is expressive of our bodily acquaintance with the world, a capacity that is grounded in the psychophysical constitution of the self as well as in its general sense of existence.

According to this model, «whatever affects me I cannot experience raw, as proto-impressions or impacts (*Uraffektion*). The very first appearance is already pervaded by affective tendencies: some form of a pre-egoic source is already affected, a world is already sketched» (Varela & Depraz 2005: 64). We respond to others' states and bodily feelings without necessarily being aware of them as objects of our intentional acts. At the same time, this form of embodied response displays what Varela & Depraz call "valence", i.e. «a dynamic polarity, as manifesting in the form of a tension that takes several forms: like-dislike, attraction-rejection, pleasure-displeasure» (Varela & Depraz 2005: 70). Valence is a form of unconscious organic fluctuation, embedded with axiological properties, and related to time consciousness. In fact, affect-

feelings are supposed to fill the "I" not like states, but rather as ramifications of the affective flow of intentionality by means of which the ego is self-organising and lives in a temporally extended present. From this point of view, basic concerns such as coping in everyday life and our habitual stance have a common structure, that is to say a feeling-tone and a complex embodied condition. Instead, affect-emotions exhibit the structure of a rainbow, in which the bodily subject is generative of valence, as well as its modes of relationship with others (Depraz 2008). According to such scheme: «The more I open up toward others, the more I am led to welcome them, to be receptive to their positive and negative emotions, to their joy or to their suffering; the more I turn toward myself, the more I contain my own feelings, be they negative (e.g., despise) or positive (e.g., admiration)» (Depraz 2008: 255). In this sense, temporality functions as the very source of the emotional rainbow.

One could say that, for Stein, empathy presupposes what Varela would have called a broader dispositional orientation towards the feelings and states of another. However, the kind of bodily self-displacement described by Stein offers a necessary but not sufficient condition for empathy. In fact, for Stein empathy involves a more complex stance towards others that includes the receptivity to spiritual formations such as values and culture. While valence is involved in the tendencies that drive us towards others, it is not a rationale of our capacity to grasp the meaning of another's state. It follows that empathy is not restricted to the experience of one's own body, for empathy includes both the capacity to respond to others' bodily states and feelings as well as to intersubjectively constituted relationships with the world and others. While Varela & Depraz hold a naturalised model that ultimately connects time dynamics with the heart-system model (Depraz 2008), Stein's account calls into question the capacity of understanding others as persons. In this sense, it seems that empathy cannot be reduced to the affective experience of the "now" in the naturalised phenomenological sense held by Varela and Depraz.

It is noteworthy that in her 1922 study on *Philosophy of Psychology and the Humanities* Stein does not take into account the problem of empathy as she rather focuses on the relationship between causality and motivation. This is particularly important if one considers that life feelings and states of living belong, in this second work, to sensate (or physical) causality. Weariness, states of excitement, exhaustion, they all determine different states of tension and manifest an enduring real property: life-power. This enduring property of the real ego, corresponding to the substrate of the sensible occurrences of mental phenomena, replaces feelings and states (Stein 2000: 24-25; Stein 2010b: 22-23). Life-power can be described as a field of generative powers that is continually operative. By contrast, motivation represents the "pointing-at" that is inherent in each intentional act. Significantly, Stein identifies motivation with apprehension, which is active, as a lower form, even in perception. Basically, Stein argues that motivation is an inclination to fulfil an intention, hence it is teleological or goal directed. As O'Conaill (2012) has put it, «in a state of being motivated, a course of action is prefigured or laid out in advance in the agent's apprehension of the motivating phenomenon. This is to say, in being motivated the agent is aware of the object *x*, and is thereby aware of and drawn towards a specific course of action, *y*. It is in this sense that an action that fits the motivated course of action is released by a meaning». Stein seems to hold this view, when she argues that motivation is a structure that is valid for the entire range of intentional experiences. However, motivation is not primarily and exclusively an inclination to act, but rather a synthesis of apprehension.

For Stein, motivation functions as "authentic spontaneity": it is what compels us to fulfil an intentional act without necessarily being a voluntary performance of the ego. It is at stake when we understand a meaning, or make judgments, but it is also responsible for the apprehension of a determinate sense-content at the level of perception. In higher levels of act-

operation, motivation is manifested as conscious, whereas in the case of perception motivation lies in the background. Furthermore, «the availability of motives does not *compel* the ego to accomplish the acts in question» (Stein 2000: 55; Stein 2010b: 48). The relevance of motivation has remarkable consequences upon the analysis of the psychophysical individual. While feelings and bodily states do not enter the connections of motivations, because they rely on life-power, hence are governed by sentient causality, higher-order feelings, such as emotions, i.e. feelings that have a value attitude (Stein 2000: 75; Stein 2010b: 65), are constituted as motives that exercise a pull. Unlike states and feelings, emotions are schematised as motives, i.e. they are not simply co-apprehended, but rather motivate me to fulfil their affective tendencies, i.e. to live them through.

One can notice that motivation does not strictly coincide with explicit reasoning, but it is rather the synthesis that takes place during an act of reasoning in that it provides continuous syntheses of temporal succession, thereby establishing continuity as well as self-coincidence. In this sense, empathic experiences always emerge towards a motive and enter into transcendental syntheses and connections of motivations. However, what enters the connections of apprehension of empathy is not the felt content, but rather the meaning content (Stein 2000: 107; Stein 2010b: 92). I can empathise with the tiredness or the sluggishness of another in virtue of nexuses of bodily acquaintance, but, when I empathize with her emotions, a synthesis of apprehension is required like in language comprehension. If I do not pay attention to a word, the physical sound alone does not express any meaning to me. Likewise, in empathy the expression of the other does not compel me to respond unless it becomes the object of my attention. For this reason, Stein stresses that we do not empathise on the basis of our own subjectivity and character: «I can experience values empathically and discover correlative levels of my person, even though my original experience has not yet presented an opportunity for their exposure» (Stein 200a: 133, Stein 1964: 104). By means of empathy, we anticipate experiences that we cannot find in our original experience. In this sense, empathy, besides drawing on bodily self-displacement, involves more complex disposition towards others that would deserve further attention.

3. Stein and Radical Empathy

As shown by Stein's analysis of the constitution of lived body, being able to decentre one's own experience provides a necessary precondition of empathy. In this sense, bodily self-displacement enables the self to respond to the feelings of another by drawing on one's own capacity to inhabit a horizon that is subjected to the shifting of the point of view. However, at the same time, empathy also requires the capacity of understanding another's mode of existence beyond perceptual features. In order to better illuminate this aspect, I shall refer to what Ratcliffe calls "radical empathy" (Ratcliffe 2012). According to Ratcliffe, the phenomenological method facilitates a distinctive kind of empathy, which he labels "radical" to emphasise that it incorporates a specific form of engagement with others. This means that, through the phenomenological stance, the world of the other enters my own world and become mine. What is distinctive of the phenomenological stance is the capacity to suspend one's own beliefs and norms in order to recognise relevant features of experience that are presupposed in our everyday life. In this sense, for Ratcliffe, radical empathy requires (i) suspending common assumptions regarding the modal space inhabited by both parties and (ii) bringing to light the modifications of that space as possibilities or changes in the sense of reality or belonging.

With regard to this, Ratcliffe takes into account Stein's phenomenology by pointing out that «her account allows for the possibility of empathic experiences that are generated by cognitive processes, rather than in perception-like way, and these might have some sophisticated contents» (Ratcliffe 2012: 476). Ratcliffe's aim is to show that there can be an empathic

experience of others' existential spaces even when these fall outside the sphere of one's own experience, at least in certain respects. For example, in the case of psychic illness or profound grief, one can refer to variations in one's own experience or make use of imagination. Such experiences can be understood through a phenomenological stance and interpreted in terms of changes in a sense of reality and belonging (Ratcliffe 2012: 486). With this regard, a striking feature of Ratcliffe's theory of radical empathy is that «the world that comes into focus through a phenomenological stance has neither a first- nor a second-person content, but appears as "ours"» (Ratcliffe 2012: 486). On Ratcliffe's account, radical empathy is a stance of openness to the other that also implies a willingness to be affected by the other. The constitution of a common world that both the empathiser and the empathised inhabit is possible only in the course of the empathic experience itself. In this sense, empathy cannot be restricted to any specific point of view, for it requires not just the translation of one's own mode of existence in that of another, but rather and more importantly the willingness to understand and interact with another's mode of existence in and through the institution of a common world.

As Ratcliffe himself notices, radical empathy is not without limits. «Once we have accepted that radical empathy is possible, there is a lot more to be said about *how* it is achieved» (Ratcliffe 2012: 487). Here, I wish to point out that both aspects of Ratcliffe's account, i.e. the appraisal of non-perceptual aspects of the empathic experience as well as the relevance of a common world made available by willingness and interaction, are central features of empathy from a Steinian perspective. For Stein, affects play a relevant role in establishing the possibility of apperceiving another as inhabiting an emotional and existential space. Yet empathy primarily consists in the possibility of making sense of the experience of another in a way that is potentially transformative and enriching for both empathiser and empathised. As Stein writes, «by empathy with differently composed personal structures we become clear on what we are not, what we are more or less than others. Thus, together with self knowledge, we also have an important aid to self evaluation. Since the experience of value is basic to our own value, at the same time as new values are acquired by empathy, our own unfamiliar values become visible» (Stein 2010a: 134; Stein 1964: 105).

This because empathy allows one to take a step beyond one's own self-experience and to participate in interactions that are regulated by laws of motivation rather than by physical causality. In engaging with the other, one needs to adopt a stance of openness, which involves spending time with the other, being affected by her, and reflecting upon the impact of the other on our own sense of reality. For this reason, Ratliffe's insights on empathy are particularly helpful because they shed light on the relevance of the empathic *attitude*. As Ratcliffe points out, empathy is not an unconscious way of simulating another's experience due to a contingent perceptual encounter. Quite to the contrary, empathy is stance of openness that establishes a potential new common world in virtue of the willingness of both subjects to participate in the empathic relationship. In relation to this, Ratcliffe refers to the transformative moment achieved within the patient-psychiatrist relation when the patient sees that her lived situation is being understood by the psychiatrist. The very fact that the doctor has an "inkling" of the patient's experience often results in a form of trust that enables and sustains the patient's process of healing. In this sense, empathy is the dimension that empowers self-knowledge without yet providing any coincidence of perceptual states between two subjects.

Conclusions

By way of conclusion, let me recall the main points of my brief analysis. I have argued that, from a Steinian perspective, empathy is grounded in a dynamic model of embodied self-experience that is partly compatible with Varela & Depraz's neurophenomenology. From this point of view, Stein's analysis of the constitution of the lived body offers important insights on bodily self-displacement. However, unlike Varela's approach, Stein holds a non-naturalised

account of lived experience. With regard to this, empathy implies attentiveness and a formed attitude to respond to the feeling of another in an appropriate way. In this sense, bodily self-displacement allows one to decentre one's own experience, but a more radical orientation towards another's sense of existence – as sustained by Ratcliffe's account – is necessary for empathy to be fulfilled.

REFERENCES

Aaltola, E. (2013), "Varieties of Empathy and Moral Agency", *Topoi*, 33, 1, pp. 1-11;
Depraz, N. (2008), "The Rainbow of Emotions: At the Crossroads of Neurobiology and Phenomenology"*, Continental Philosophy Review,* 41(2), pp. 237–259;
Dullstein, M. (2013), "Direct Perception and Simulation. Stein's Account of Empathy", *Review of Philosophy and Psychology*, 4, pp. 333-350;
Geiger, M. (1911), Über *das Wesen und die Bedeutung der Einfühlung*, Barth, Leipzig;
Haney, K. M. (2013), "Edith Stein and Autism", In Embree, L. and Nenon, T. (eds.), *Husserl's Ideen*, Springer, Dordrecht, pp. 35-54;
Lebech, M. (2015), *The Philosophy of Edith Stein. From Phenomenology to Metaphysics*, Peter Lang, Bern;
Meneses, R. W. and Larkin, M. (2012), "Edith Stein and the Contemporary Psychological Study of Empathy", *Journal of Phenomenological Psychology*, 43, pp. 151–184;
O'Conaill, D. (2013), "On Being Motivated", *Phenomenology and the Cognitive Sciences,* 12(4), pp. 579–595;
Pinotti, A. (2000), "Moritz Geiger fra analogia e differenza. Note sul metodo di un «fenomenologo solo per un quarto»", In Besoli, S., and Guidetti, L. (eds.), *Il realismo fenomenologico. Sulla filosofia dei circoli di Monaco e Gottinga,* Quodlibet, Macerata, pp. 445-466;
Ratcliffe, M. (2012), "Phenomenology as a Form of Empathy", *Inquiry*, Vol. 55, No. 5, pp. 473–495;
Sheets-Johnstone, M. (2015), "Embodiment on Trial: A Phenomenological Investigation", *Continental Philosophy Review,* 48, pp. 23-39;
Stein, E. (1964), *On the Problem of Empathy*, translated by W. Stein, Springer, Dordrecht;
Stein, E. (2000), *Philosophy of Psychology and the Humanities*, translated by Baseheart, M. C., and Sawicki, M., ICS Publications, Washington D.C;
Stein, E. (2010a), *Zum Problem der Einfühlung*, edited by Sondermann, M. A., Herder, Freiburg;
Stein, E. (2010b), *Beiträge zur philosophischen Begründung der Psychologie und der Geisteswissenschaften*, edited by Beckmann-Zöller, B., Herder, Freiburg;
Svenaeus, F. (2015), "The Phenomenology of Empathy: A Steinian Emotional Account", *Phenomenology and the Cognitive Sciences*, 15(2), pp. 227–245;
Szanto, T. and Moran, D. (2015), "Introduction: Empathy and Collective Intentionality – The Social Philosophy of Edith Stein", *Human Studies. A Journal for Philosophy and the Social Sciences*. Stein's special Issue, 38(4), pp. 445–461;
Varela, F. J. (2000), "The Specious Present: A Neurophenomenology of Time Consciousness", In Petitot J., Varela F. J., Pachoud B., and Roy J.-M. (eds), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, Stanford University Press, Stanford, pp. 266-329;
Varela, F. J. and Depraz, N. (2005), "At the Source of Time. Valence and the Constitutional Dynamics of Affect", *Journal of Consciousness Studies*, 12, No. 8–10, pp. 61–81;
Vendrell Ferran, I. (2015), "Empathy, Emotional Sharing and Feelings in Stein's Early Work", *Human Studies. A Journal for Philosophy and the Social Sciences*. Stein's special Issue, 38(4), pp. 481–502;
Zahavi, D. (2014), *Self and Other. Exploring Subjectivity, Empathy and Shame*, Oxford University Press, Oxford.

SANDY BERKOVSKI
Bilkent University
sandy.berkovski@gmail.com

A NATURALIST VIEW OF HUMILIATION[1]

abstract

A naturalist analysis of humiliation begins with the notion of social interaction, a public encounter with other people. Interactions are an essential element in cooperation, a vital condition of survival and well-being. In the course of interaction the person presents himself as someone possessing the qualities necessary for successful cooperation. An act of humiliation is designed to inflict damage on the agent's self-presentation. Any such damage would be a sign that the agent is not successful in conducting the given interaction. Such damage would tend to decrease of cooperative value of the humiliated individual and to decrease his chances of survival and reproduction.

keywords

naturalism, humiliation, embarrassment, social pain, humour

Phenomenology and Mind, n. 11 - 2016, pp. 78-86
DOI: 10.13128/Phe_Mi-20108
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

It is prohibited to humiliate a person. A humiliating, degrading treatment of an individual by another individual, or an institution, constitutes a serious moral transgression. The task is to understand the nature of humiliation, to explicate its concept, and to discover the reasons behind the prohibition. Some theorists writing on this topic favour a broadly Kantian solution[2]. Each human being, in virtue of his capacity to act rationally and the related capacity to act morally, has an intrinsic worth. The purpose of an act of humiliation is to deny him his worth. Every humiliating act is literally inhumane: in humiliating someone we treat him as a non-human—a beast perhaps, or an inanimate object. Every action directed toward another individual, including a punishing action, must include a minimal doze of respect, a respect of him as a human being.

This solution faces two major difficulties. (1) Not every act of humiliation destroys humanity, or can even be plausibly said to attack humanity, i.e. either humanity as a universal or just the human nature of the particular individual. Taunting or ridiculing someone is a typical case of humiliation, but it is not at all clear why any such act should be taken to be an assault on humanity. While insults are often humiliating, can they really represent an assault on human dignity? (2) Being humiliated is a painful experience (in some sense of 'pain', more on which below). It seems, however, that the Kantian account does not assign any explanatory role to the psychological suffering of the victims of humiliation. It seems that, even if the victim were *not* to suffer, the act of humiliation, according to the Kantian, would be just as illegitimate. The perpetrator would still be mounting an assault on the victim's humanity, its success in no way correlated with the victim's psychological experience. Suffering of the victim, in other words, is not a ground for condemning humiliation. On the other hand, one might add, if suffering should play any role in the prohibition of humiliation, the Kantian account does not attempt to explain the source of that suffering.

These difficulties are not insurmountable. They would not be sufficient on their own to dismiss the Kantian and semi-Kantian accounts of dignity. Still, they are sufficient to motivate a search for an alternative interpretation. The broadly naturalistic approach I pursue here builds on two general premises. One is methodological: ethical reasoning should be continuous with the reasoning in natural sciences. It should necessarily take into account what natural sciences

1 I am grateful to the anonymous referee for helpful comments on the earlier draft of this paper.

2 See Darwall (2006:119–150) and Margalit (1996:57–75).

tell us about the behaviour and constitution of humans. As such, it follows the findings of contemporary biology and empirical psychology. The second assumption is substantive. The naturalistic approach sees a human individual as wholly a part of nature, and specifically, of the animal kingdom. Man is subject to the forces of natural selection. To understand human behaviour we have to look at it through the lens of evolution. The institution of morality, moral judgements, norms, and intuitions, can only be made intelligible if we locate them within the evolutionary process. The model I shall propose contrasts humiliation with shame and aligns it with embarrassment. Acts of humiliation aim at damaging and destroying the victim's cooperative value, thus severely hampering his chances of survival and reproduction. Pain is an appropriate reaction to such acts. Finally, the moral ban on humiliation should be justified entirely by reference to the pain it causes.

2. Shame, humiliation, embarrassment

Many writers, from Hobbes onward, noted our exceptional sensitivity to slights, rebukes, insults by others. Various explanations have been suggested, but the path to a correct one, I think, goes through a distinction between two kinds of occasions on which we display such sensitivity. Suppose, for instance, I find a diary my grandfather kept for the last ten years before his death. In it there are reflections on my life and my character. In the course of his thoughtful remarks the grandfather comes to the conclusion that I am lazy and dishonest. Since I think of myself as honest and hard-working, I am understandably upset. As I value my grandfather's judgement, I am prompted to examine my own character. Eventually I realise how right he was, and I resolve to change myself to the better.

In a rather different scenario, suppose that I find myself together with my uncle among distant acquaintances with whom we occasionally play tennis. I have sufficient respect for my uncle; I sufficiently value the mental and moral qualities of our tennis partners. As the conversation takes an unusual turn, my uncle begins questioning my integrity, as well as my working habits. I react with anger, and we nearly come to blows. I vow to never talk to him again.

When all the variables are taken into account, the evident difference between the two scenarios is that the former one unfolds in private. It is not far-fetched to think that the grandfather's diary caused damage to my self-esteem. An emotional reaction to that damage was *shame*. And because I was ashamed of my character flaws, I initiated the process of self-improvement. Shame is thus aligned with reflection and self-examination.

A very similar criticism on a public occasion yielded rather different results. Here we want to say that my uncle caused my *humiliation*: I was humiliated in front of those acquaintances. In fact I might have been humiliated even if there were no observers around, other than the uncle himself. No self-examination accompanied my experience, or followed it later on. My emotional response was immediate and instinctive. Even if there were any such self-examination, I might have found no fault with myself. My self-esteem would have been intact.

Perhaps, however, you are not convinced. Perhaps you want to say that I was *embarrassed*, rather than humiliated. To embarrass someone is a minor offence. The experience of embarrassment is mildly distressing. It is indeed tied to a social occasion. 'Humiliation' should be reserved for most serious assaults on my very humanity. It is only incidentally tied to a social occasion, if at all.

Some might find this elevated view of humiliation compelling[3]. Instead of pitting my intuitions against theirs, let me follow their lead and first give a naturalistic analysis of embarrassment. I will attempt to show that, once properly interpreted, embarrassment should be seen as being on a par with humiliation. Situations of embarrassment and humiliation are of the same kind.

3 See Margalit (1996).

Both are situations of social exclusion. The only significant difference is the scale—of offence, of suffering, and of lasting damage.

3. Self-presentation

To understand, therefore, the phenomenon of embarrassment, begin with the notion of social interaction. Conversation is a paradigmatic instance of interaction, yet we cast the net much more widely. Exchanging glances on the street, riding on a bus—all of these are also its instances. Being merely in the presence of others, where your presence is routinely observed, would be the fundamental unit of interaction. Every participant in such interaction comes equipped with a set of expectations of how he should behave in it and, on the other hand, how others are supposed to behave as well. Through his behaviour, passive or active, he presents himself to others. Not only does he have his own expectations of his own self-presentation, but also he has an idea of what expectations others have formed of *his* self-presentation.

The object of embarrassment is the person's self-presentation[4]. This elusive concept can be explicated as a way of carrying myself in public. I present myself as someone deserving a certain treatment from others in accordance with my expectations. I expect and demand that treatment from others. Let us say that, while in public, I project a 'self-presenting image'. A major concern of the interaction participant is to protect and maintain the self-presenting image. So it is not only the actual physical treatment I receive from others (words, gestures, physical contact) that is the focus of my concern, but also the opinions that others may have formed of me. Only rarely is my construction of that image deliberate and planned. On most occasions, it is habitual and instinctive, and its parameters are not clearly articulated. Sometimes I present myself as an intelligent individual, sometimes as a funny one, and often as a courteous one. But without exception I want to be seen as someone capable of maintaining the interaction. I want to be in control of my impulses, and I loathe to be the one causing the smooth flow of interaction to cease. When in this fashion I cause an interruption, my self-presenting image is damaged. By the same token, I can embarrass someone else—which is another way of saying that I can damage someone else's self-presenting image.

Why embarrassment, being a reaction to the damage of self-presenting image, should be painful is hardly mysterious. Social interactions serve to initiate, maintain, and terminate our relationships with other people. The person's qualities are manifested in those interactions—if not exclusively, then primarily so. It is essential for the well-being of any person to have good relationships with people. Even if we imagine a person materially independent of other people, it is not possible for anyone to be emotionally completely independent of them. Interaction, then, is a essential element in cooperation, a vital condition of survival and well-being.

Now, a damage to the agent's self-presenting image is a sign that the agent is not successful in conducting the interaction. A mild such damage does not indeed rule him out as a future partner in interaction, but it places a question mark over his capacity all the same. Slipping on a street is not sufficient for casting you outside of the society. It is nevertheless a sign of clumsiness, or perhaps of ill health. Neither of these qualities would increase your value in future cooperative activities. Beyond this sort of slightest damage, this is clearer in a conversation when a silly remark, or a tasteless joke, throw a conversation off-balance. No one knows how to react, and the author of the remark is no longer seen, at least at that moment, as belonging to the group involved in the interaction.

4 The concept of self-presentation is derived from Velleman (2006a). However, I draw differently the lines connecting it with shame, embarrassment, and self-esteem. My understanding of embarrassment is heavily influenced by Goffman (1967c) and Goffman (1967a). For other views on the nature of embarrassment see Keltner and Buswell (1997).

4. Physical pain and social pain

'Being in an embarrassing situation is painful.' Commonsensical as it might sound, this statement raises two questions. Is the talk of pain here metaphorical at best, or are we talking of real, literal pain? And secondly, whatever the status of pain mentioned here is, how does it fit into the explanatory pattern outlined above? It is tempting to think that pain occasioned by embarrassment and pain occasioned, say, by burning are different physiological processes. The former is simply a variation of sadness, a particular mood. The latter is pain properly speaking, associated with physical injury.

Recent research undermines the possibility of any such distinction. Begin with the function of physical pain: why should any such capacity for the subjective experience of pain have evolved? It is tempting to think of pain as an indicator of a problem. If you are injured, pain allows you to pinpoint the location of injury that is the source of pain. But this seems false at least in the case of internal pain. If you have gastrointestinal pain, you will not be able to find its source merely by introspection. The experience of pain cannot direct you toward its own source. The capacity for experiencing physical pain has not evolved to enable us to identify the source of malfunction (e.g., a damaged organ).

More promising is to think of pain as an evolutionary instrument for alerting us to a problem in our body and for initiating a sequence of behavioural reactions[5]. Escape from, or removal of, possibly harmful (pain-producing) stimulations, inhibition of activities that might delay recovery are among them. Its other role is in avoiding potentially damaging situations (and showing others the effects of those situations). It is not merely that, without a pain experience, I would not be able to tell that my hand is burning, with the resulting damage to tissue and bone. More importantly, I would not necessarily avoid the damage of tissue and bone: after all, it requires some non-trivial competence to understand how this damage threatens the survival of the organism.

The pain of embarrassment, the subjective experience we provisionally call 'pain', can be interpreted in a similar way. A damage to cooperative value increases chances of social exclusion. It thus restricts our access to material resources, with the ensuing threat to our survival. The pain of embarrassment serves as a signal that something went, or is going, wrong in the social interaction. Without the pain of embarrassment we would not have known of the damage to our self-presenting image. As in the case of physical pain, the pain of embarrassment ensures that we will try to avoid this behaviour on other occasions. And in addition, a manifestation of embarrassment teaches others as well that such behaviour should be avoided.

There is, in other words, a nearly perfect match in the ultimate explanations of physical pain and of the pain of embarrassment. Is there a match in their proximate explanations that would cite the underlying psychological or physiological factors of particular experiences? Without taking a far-reaching stand in the philosophy of mind, the question is whether the same psychological and physiological processes can be associated with the two subjective experiences of pain. Embarrassing situations are situations of social exclusion of the embarrassed individual: the failure to protect one's self-presenting image diminishes the cooperative value of that individual. The individual is facing a threat of being excluded from the group involved in the given interaction. With this assumption granted, embarrassment pain becomes what the psychologists call 'social pain', a subjective experience accompanying social exclusion. Recent studies demonstrated a significant overlap, at the level of brain circuitry, between socially painful and physically painful experiences[6]. The same regions of

5 See Bateson (1991:829), Dennis and Melzack (1983:153ff), and Wall (1979:257).

6 See Eisenberger (2012).

the brain are activated in the two types of experience. Further remarkable correlations have been demonstrated. Factors increasing physical pain have been found to increase social pain. Factors increasing social pain also increase physical pain. The same two correlations hold for the decrease of the relevant kinds of pain.

The precise correlations are still a matter of active research. A philosophical conclusion for us to draw at this stage should be this. Embarrassment is accompanied by a kind of pain that is no less real than physical pain, the kind of pain accompanying damage to tissue and bones. A normative implication follows: a deliberate attempt to embarrass a person can be prohibited on the same grounds as a deliberate attempt to injure a person.

5. Humiliation reconsidered

Embarrassment, I have now claimed, is a response to the situation of a failure to meet the expectations of the participants in a social interaction, as understood by the embarrassed agent himself. Its proximate cause is the concern for self-presentation in the given interaction. Its ultimate cause is the decrease of cooperative value. The pain of embarrassment is a form of social pain.

We can now substantiate the claim advertised in §2: humiliation is embarrassment writ large. Both involve damage to self-presentation. An act of humiliation is a deliberate attempt to damage a person's self-presenting image. Of course sometimes we say that a person humiliated himself, just as we might say that a person embarrassed himself. In both cases the person is understood to have unintentionally humiliated or embarrassed himself. And so, properly speaking, there is no action involved, but merely a series of accidents creating a situation of humiliation. This situation is characterised by a damage caused to the self-presenting image of the given person. The emotion of humiliation is the response of the humiliated person to the situation of humiliation.

The extreme ease with which we can be humiliated is parallel to the ease with which we can be embarrassed, or embarrass ourselves. At stake is our self-presenting image and, in consequence, our cooperative value. Humiliation and embarrassment are phenomena of the same kind differing only in intensity. An extensive damage to self-presenting image is humiliation. An intense pain felt in response to it is the pain of humiliation. However, are there any substantive reasons we can give for this conceptual identification of embarrassment and humiliation? The vernacular that often merges the two terms and their cognates is not a reliable guide. Introspection fares no better: occurrent mental states are too similar to tell them apart[7].

There are, I think, two ways of showing embarrassment and humiliation to be of the same kind. One is to contrast both with shame, another emotion often conflated with them. As the example in §2 has shown, shame is a response to a drop of self-esteem. It is an emotion aligned with reflection. As I reflect on my shortcomings displayed in a given situation, my self-esteem drops. Shame follows as an emotional response. Shame, on this view, will be a reflective emotion, an outcome of a reflective self-examination. In recalling my past behaviour, or even my thoughts, I can realise how wrong they were, how I failed to live up to the ideals constituting my self-esteem. No presence of an actual audience is required. This is not possible with either humiliation or embarrassment: I cannot experience these emotions in solitude[8].

7 Situations of extreme public humiliation do elicit strong feelings, situations of average embarrassment elicit weaker ones. It remains unclear, I think, whether the difference is of quality, not merely of degree. This is confirmed by our linguistic intuitions. On many occasions, 'He humiliated me' can be substituted, without sounding inappropriate, for 'He embarrassed me', and *vice versa*. If the feelings were clearly distinct, these substitutions could not easily pass.

8 The possibility of solitary shame was forcefully defended in Williams (1993).

Moreover, these emotions are emphatically not the products of self-evaluation. They are immediate responses to certain kinds of situations, namely, situations liable to decrease one's cooperative value.

The second reason for aligning embarrassment with humiliation is the unique role of ridicule in both of them. Flipping on the street is often funny—so much so that polite observers are required to contain their laughter in order not to embarrass the person on the ground. Alas, he is already embarrassed, in part because he realises how ridiculous he looks. In Agatha Christie's novel, Lady Edgware misunderstands the expression 'judgement of Paris' taking 'Paris' to refer to the city and its world of fashion, rather than to the son of Priam. It is a *faux pas* creating embarrassment, but also liable to provoke laughter among the less reserved onlookers. Her Ladyship's boorishness is a thing to poke fun at. If laughter is elicited and encouraged, we are in the terrain of humiliation. Ridicule is its most potent weapon.

Why should that be so? In line with the present methodology, we ought to understand the evolutionary function of humour. There is a whole array of competing accounts available[9]. They fall into two broad categories. Some highlight the competitive and somewhat menacing side of humour. Others insist on the continuity between humour and play, and on its softer, pleasant aspects. As a representative of the first category of views, consider the ostracism theory due to Richard Alexander[10]. The capacity for laughter and the trait enabling us to appreciate humour, according to it, have evolved to decrease the social status of the object of a joke (to take one instance of a humorous activity). It simultaneously aimed to increase or reinforce the social status of the joke teller and of his audience. Both capacities originate in grooming and tickling among primates. These latter activities were from the beginning designed to establish and maintain bonding between individuals: grooming requires time and effort. One necessarily has to be selective in choosing one's grooming partners. Bonding is thus necessarily exclusive. Humour as we know it inherits these ancient features. It is supposed to enhance the social bonds among the individuals sharing, e.g., a joke. It simultaneously is supposed to exclude, or 'ostracise', the individuals who are the objects of the jokes and of other humorous activities, such as puns.

A representative of the second category is a view advanced by Gervais and Wilson[11]. Essential to these authors' approach is the distinction between Duchenne and non-Duchenne laughter. The fundamental trait that modern humans inherited from early hominids is the ability to react to a certain a type of stimuli and to experience positive emotions. This is a characteristic of Duchenne laughter. It is genetically transmitted: infants are able to laugh without being taught to do so. The evolutionary origins of Duchenne laughter, as in Alexander's theory, are located in the primates' activities of grooming and tickling. The ultimate evolutionary cause of Duchenne laughter, however, is not ostracism and bonding, but rather play and unexpected social situations that are not life-threatening. It is of course unclear how play can fail to have a bonding function. And if it is bonding, then this should come at the expense of bonding with others, and hence necessarily involve ostracism. The latter notion should be weak enough, as indeed it is according to Alexander, to involve a mere *de facto* exclusion of individuals from the given activity by the group. It is in this sense not a hostile act. If, on the other hand, we curtail the notion of ostracism to include only deliberate, hostile acts aimed at preventing individuals from being part of the group, then Duchenne laughter can no longer accompany ostracism.

9 See, e.g., Alexander (1986), Weisfeld (1993), Gruner (1979), Gervais and Wilson (2005), Provine (2012). For a critical review see Martin (2007).

10 See Alexander (1986).

11 See Gervais and Wilson (2005).

A laughter accompanying hostile acts of ostracism is not a direct response to stimulus and is not associated with that positive emotion alternatively termed 'mirth' or 'joy'. The hostile, deliberate act of ostracism may involve a strategic deployment of laughter—i.e. a different kind of it, a non-Duchenne laughter.
According to both approaches, therefore, humour is a major instrument of social exclusion. This is due either to the fact that the ultimate cause in the development of humour consisted in creating such an instrument, or to the fact that humour's ultimate cause (play) had an incidental feature which could be exploited to promote social exclusion. Being made fun of is painful because it is a sign of one's alienation from a group.

6. Normative implications

We have now isolated a cluster of the situations of social exclusion and the associated emotional responses. Contained in it are insult, ridicule, embarrassment, and humiliation. Experiences of the subjects in these situations is quite literally painful. There is, as we have seen, more than one way to draw an analogy with physical pain. Similarities are observed or inferred in behavioural patterns, evolutionary explanations, and neurochemical responses. Based on these findings, there is a straightforward suggestion to make. Deliberate acts of exclusion are acts of torture, a deliberate infliction of pain. They are painful to their victims in much the same way that acts of physical torture are. Thus, the prohibition on humiliation should be based on the prohibition on inflicting unnecessary pain. If the boundary between social pain and physical pain can be drawn only arbitrarily, then infliction of social pain can only be as permissible as infliction of physical pain. That much is warranted by a naturalist analysis of the exclusion cluster and of the emotions associated with it.

REFERENCES

Alexander, R. D. (1986), "The reproductive significance of humor", *Ethology and sociobiology*, 7, pp. 253–270;
Bateson, P. (1991), "Assessment of pain in animals", *Animal Behaviour*, 42, pp. 827–839;
Darwall, S. L. (2006), *The Second-Person Standpoint*. Harvard University Press, Cambridge MA;
Dennis, S. J. and Melzack, R. (1983), "Perspectives on phylogenetic evolution of pain expression", In Kitchell, R. L. and Erickson, H. H., (Eds.) (1983), *Animal Pain*, Springer, Berlin, pp. 151-160;
Eisenberger, N. I. (2012), "The neural bases of social pain", *Psychosomatic Medicine*, 74(2), pp. 126–135;
Gervais, M. and Wilson, D. S. (2005), "The evolution and functions of laughter and humor", *The Quarterly Review of Biology*, 80(4), pp. 395–430;
Goffman, E. (1967a), "Embarrassment and social organization", In Goffman (1967b), *Interaction Ritual*, Pantheon Books, New York, pp. 97-112;
Goffman, E. (1967b), *Interaction Ritual*, Pantheon Books, New York;
Goffman, E. (1967c), "On face-work". In Goffman (1967b), *Interaction Ritual*, Pantheon Books, New York, pp. 5-45;
Gruner, C. R. (1979), *Understanding Laughter*, Rowman & Littlefield, New York;
Keltner, D. and Buswell, B. N. (1997), "Embarrassment: its distinct form and appeasement functions", *Psychological Bulletin*, 122(3), pp. 250–270;
Kitchell, R. L. and Erickson, H. H., (eds.) (1983), *Animal Pain*, Springer, Berlin;
Margalit, A. (1996), *The Decent Society*, Harvard University Press, Cambridge MA;
Martin, R. A. (2007), *The Psychology of Humor*, Elsevier Academic Press, Amsterdam;
Provine, R. R. (2012), *Curious Behaviour*, Harvard University Press, Cambridge MA;
Velleman, J. D. (2006a), "The genesis of shame", In Velleman (2006b), *Self to Self*, Cambridge University Press, Cambridge, pp. 45-69;

Velleman, J. D. (2006b), *Self to Self*, Cambridge University Press, Cambridge;
Wall, P. D. (1979), "On the relation of injury to pain", *Pain*, 6, pp. 253–264;
Weisfeld, G. E. (1993), "The adaptive value of humor and laughter", *Ethology and sociobiology*, 14, pp. 141–169;
Williams, B. A. O. (1993), *Shame and Necessity*, University of California Press, Berkeley.

SESSION

SESSION 2

EMOTIONS, MORALITY, AND POLITICAL THEORY

ANNA ELISABETTA GALEOTTI
Università del Piemonte Orientale, Vercelli
elisabetta.galeotti@uniupo.it

STRAIGHT AND TWISTED SELF-DECEPTION

abstract

The paper analyzes the two types of self-deception, usually labeled straight and twisted self-deception. In straight cases the self-deceptive belief coincides with the subject's desire. In twisted cases, by contrast, the self-deceptive belief opposes the subject's desire as in the example of Othello's conviction of Desdemona's infidelity. Are both these contrasting types of deceptive beliefs cases of SD? The argument of this paper shall answer this question in the positive, yet in different way from the unitary explanation of straight and twisted SD proposed by Alfred Mele. The causal account of SD claims to provide a unitary and simple explanation for both straight and twisted SD, and considers such a unitary explanation as a specific virtue of the causal view. Within the same causal model, the difference between straight and twisted self-deception is explained by a difference in the motivational state that in twisted cases is dominated by emotions. The paper will critically examine this claim, and advance an alternative explanation based on a different view of self-deception where emotions play a role alongside wishes both in straight and in twisted case.

keywords

self-deception, straight and twisted, causal account, desire and emotions

Phenomenology and Mind, n. 11 - 2016, pp. 90-99
DOI: 10.13128/Phe_Mi-20109
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introductory remarks

In this paper, I intend to face the issue of two types of self-deception (SD) usually labeled straight and twisted self-deception. I take that SD is believing that P against the available evidence and on the basis of one's wish that P. In straight cases, the self-deceptive belief coincides with the subject's desire. For example Swann, the hero of the first book of Proust's *La Recherche,* desperately wishes that his lover Odette is faithful to him; hence he discounts all the evidence suggesting Odette's infidelity so as to go on believing her faithful[1]. In twisted cases, by contrast, the self-deceptive belief opposes the subject's desire as in the example of Othello who, despite wanting Desdemona faithful to him, becomes convinced of Desdemona's infidelity in the absence of the appropriate evidence. Can we consider both these contrasting types of deceptive beliefs as cases of SD?

In the philosophical discussion, SD is accounted for by two opposite models: the intentional, according to which SD is a doing of the subject, though non-transparently, and the causal, according to which SD is a mental event caused by a motivational state triggering cognitive biases. Twisted SD represents one of the several weaknesses of the intentional model, as noted by the eminent representative of the causal account, Alfred Mele (2001). Intentionalists view SD as a non-transparent plan of the subject, vicariously serving her goal to believe what she wishes to be the case (Davidson 1985, Talbott 1995, Bermudez 2000). In twisted cases, however, the subject ends up believing what she does not wish, hence the purposive nature of her deceptive belief is undermined. Why should anyone engage in such a self-sabotaging plan, though unconsciously? Within the intentional account, twisted SD is a problem, requiring a different kind of account.

By contrast, the deflationary causal-motivationalist account of SD claims to provide a unitary and simple explanation for both straight and twisted SD, and considers such a unitary explanation as a specific virtue of the causal view (Lazar 1997, 1998; Mele 2001, Scott-Kakures 2002).

I will critically examine this claim, and argue that the unitary model provided by the causal account pays the price of its deflationary approach and is much less unitary than is claimed. I shall propose an alternative explanation, which makes sense of the difference between straight and twisted cases; my alternative explanation is based on what I have labeled the

1 Both Swann and Othello have more complicated attitudes, feelings and beliefs, but here I consider them only in sketchy manner as types of straight and twisted SD

'Invisible Hand model of SD' (Galeotti 2012) and is carried out by means of a more detailed analysis of the composition of the motivational state.
My argument proceeds as follows: I shall start by presenting the deflationary account as provided by Alfred Mele (2001), and pointing out its weaknesses. Then, I shall briefly outline my alternative account of SD, patterned after the invisible hand model, and turn to the unitary explanation of straight and twisted cases. I shall focus my critical analysis on unpacking the black box of the "motivational state." In this way, I shall be able to understand the similar mechanism in both straight and twisted cases as well as to explain their difference. The difference between the two SD types lies in the different emotional response to the threat of P not being true, each leading to a specific cognitive option and to related biases.

2. The deflationary account critically examined

Mele lists the following as jointly sufficient conditions for SD:

1) the belief that *p* which S acquires is false;
2) S treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way;
3) This biased treatment is a non-deviant cause of S's acquiring the belief that *p*,
4) The body of data possessed by S at the time provides greater warrant for *non-p* than for *p*.

I hold that these four conditions are actually insufficient a) to characterize SD as a specific phenomenon in the range of motivated irrationality; b) to explain the selective working of SD, which is not the general cognitive shortcut taken all the time we meet with unwelcome evidence; c) to provide a satisfactory unitary model for both straight and twisted cases. My argument will concentrate on the third issue, but will advance some remarks also on the others.
The four sufficient conditions state that a belief is self-deceptive if 1) it is false; 2) it is causally produced by cognitive biases; 3) triggered by a (yet undefined) motivational state; 4) when the subject possesses evidence contrary to the belief in question. Condition (4) does not specify if the data "possessed by S" are merely available to her, in front of her eyes, or are actually appraised by S. Moreover, the order of the four conditions is such that the intrinsic link with condition (2), concerning the biased treatment of data, is obscured. Actually SD is initiated because S has appraised data contrary to, hence undermining, the truth of S's wish that P. S's emotional response to this threat leads her to linger on her thoughts which become easily affected by biases, ending up in the belief that P, which is false but corresponds to S's anxious wish.
The two points of a) the appraisal of contrary evidence and b) the consequent priming of the biased treatment of data, and in that order, are to my mind crucial to account for SD as a specific instance of motivated irrationality. If the subject believes what she wishes, and if it contingently turns out that what she wishes happens to run against available data, then SD is just a case of wishful thinking gone badly. If instead SD is the motivationally distorted response to the threat of the negative evidence that the subject perceives or suspects as running against her wish that P, then it is a distinct phenomenon in the range of motivated irrationality. In the first case, SD is not sufficiently marked off doxastic states such as wishful-thinking and illusions. The distinctiveness of SD is simply entrusted to condition (1), that P is false, which however is just definitional, and does not affect (2), that is the production of the self-deceptive belief. In order to preserve the specificity of SD I hold *firstly* that condition (4) must be revised as "S appraises of data contrary to her wish that P, making her suspect that P is under threat"; and *secondly* the ordering position of the four conditions must be changed to make clear that the biased treatment of data is primed by the threatening suspicion aroused by the appraisal of negative data. Moreover, the distinctiveness of SD further requires that the condition "S has a wish that P" be added to the list. A distinctive feature of self-deceptive

beliefs compared to other motivated false beliefs, such as stubborn ones, is that the content of self-deceptive belief is thematically linked to the motivating wish (Lynch 2013).
If the SD process starts as the reconsideration of the negative evidence concerning the wish that P by a subject under a motivational state, then the process cannot simply be described as the causal triggering of motivation on biases which in turn causally affect the subject's reasoning. I agree that the product of SD is unintended, but I hold that the mental process producing SD is intentionally entered into as a response to the suspicion of ~P. Thinking matters over, when a threat to one's own crucial wish that P, is not epistemically faulty; but the emotionally loaded wish that P is likely to unleash cognitive biases in considering the available evidence, bringing about the self-deceptive belief. Briefly, I propose an explanation of SD patterned after the 'invisible hand model' to explain SD. The model, advanced for the explanation of seemingly purposeful social phenomena (Nozick 1977), provides a neat account of the coexistence of intentional doing, causal composition, and unintentional outcome, for it explains a certain beneficial effect as "the result of human action, but not of human design" (Hayek 1952). Similarly, in SD the intentionality of process must be sharply distinguished from the non-intentionality of outcome, despite the apparently self-serving result. The invisible hand model can also make sense of the apparent purposefulness of the SD product, which, though it is not a goal of the self-deceiver, however is not simply contingent. Arriving at the false belief that P is the effect of the subject's biasing, but stopping at P is her intentional doing. As I shall show late, also twisted cases, in a way, serve a goal of the subject though definitely not brought about with that intention.

3. Desires in Straight and Twisted Cases

Turning now to straight and twisted case, the analysis of motivation, as suggested by the causal model, raises the following controversial issues:

1) Does the content of the motivational state change in the case of twisted SD compared to straight SD? If it does, is the explanation of both straight and twisted cases still unified? (Mele 2001, Lazar 1999, Scott-Kakures 2002).
2) Is the operative desire for producing either straight or twisted SD the desire that P be the case or the desire to believe that P? (Nelkin 2002)
3) Should there be a match between the operative desire and the content of self-deceptive belief or is the operative desire content-unrestricted compared to the self-deceptive belief? (Nelkin 2002, 2012).

Concerning point (3), I agree with Nelkin that there should be a match between the operative desire and the content of the self-deceptive belief, otherwise a) it is not clear that it is a case of SD, for the self-deceptive belief must be thematically linked with the original motivation; b) methodologically, the shift looks an ad hoc adjustment. Think of the example of Swann: his self-deceptive belief that Odette is faithful to him corresponds to his desire. Such correspondence allows any external observer to suspect SD, given that the resulting belief that P fulfills Swann's desire and, at the same time, runs against the evidence available to him. Next, consider the example of Othello. Following Mele, Othello's original desire of Desdemona's fidelity is not the operative one producing the (false) adverse belief that she is guilty. According to Mele, here the original desire is trumped by another—the desire of not being fooled—which becomes operative in triggering data biasing and producing the false belief of Desdemona's infidelity. Thus, the operative desire has no content-match with the resulting self-deceptive belief, and in Nelkin's terms it is content-unrestricted.
Point (2) is raised by Dana Nelkin in her attempt to find an alternative solution to the content-restrictedness issue for twisted SD. She thinks that it is possible to keep unitary the causal explanation in both straight and twisted case if what the subject wishes is not P, but rather *to believe that P.* Swann wishes *to believe Odette faithful*, in order to keep his peace of mind, while,

symmetrically, Othello wishes *to believe Desdemona unfaithful*, in order to avoid being fooled. Under this description, in both cases there is a match between the operative desire and the self-deceptive belief. This apparently neat solution however leaves much unanswered. Does the self-deceiver wish to believe that P, no matter what, or rather wish that the belief that P be true and consequently be believed? Why should anyone want to believe that P no matter what? The self-deceiver is epistemically confused, but is not incoherent or crazy. Those who think that the operative desire is *to believe that P* assume that there has been an unconscious trade-off between the wish that P be the case and the wish to believe P no matter what. True, at the end of the self-deceptive process, the subject precisely gets the belief that P devoid of its truth-value. But what the subject eventually gets does not make such belief the primary object of her desire, nor of her unconscious aim. Actually the supposed trade-off between states-of-the world and beliefs would make the pursuit of the false belief that P a crazy intent indeed (Lazar 1996). Nelkin's proposed solution for keeping unitary the explanation of straight as well as twisted cases must be rejected.

Coming now to point (1) concerning the composition of the motivational state, Mele holds that there is a difference between either straight or twisted SD (2002, 111ff). In the former, the driving component is desire, in the latter is emotions; other scholars supporting the causal account instead point to different types of emotions in either SD types (Lazar 1999). In general, in the causal account, the specific role of emotions and desires is not clearly set apart. Mele, on the one hand, attributes the difference between straight and twisted cases to the different causal role of desires and emotions respectively; on the other hand, illustrating the case of the jealous husband, he also singles out an operative desire triggering the biasing as well, the one criticized by Nelkin for not being content-restricted. Ultimately, it is not clear whether the twisted cases are caused by emotions or by the operative desire or by both, being the emotions the cause of the desire's switch. The lack of a clear specification of the motivational state weakens the causal account of the twisted cases, and, besides, attributing the causal factor in one case to desires and in the other to emotions makes the account less unitary than claimed.

I argue instead that the components of the motivational state are the same in both straight and twisted SD, and, more precisely, that the motivational state giving rise to SD processes of all sorts comprises the desire that P and the emotions of fear and anxiety loading the desire that P and making it "rebellious" (Pears 1994). How to account of such composition of the motivational state? Given the desire that P—often already present, albeit dormant in the subject—two elements contribute to engender the motivational state favorable to start a self-deceptive belief formation process. *The first* is the appraisal of the evidence threatening the truth of P. Before that moment, Swann or Othello did not think of their lovers' fidelity or infidelity. It is only when Odette fails to meet Swann at the Verdurin's salon one night, or when Iago insinuates doubts about Desdemona, that jealousy is aroused. Since the emotional arousal proceeds from a pre-attentive appraisal and not from a proper epistemic attendance to and processing of, the negative evidence, it may well be the case that such appraisal is actually groundless. This is important for the account of twisted SD, such as Othello-like jealousy. In any case, the suspicion about one's lover's fidelity arouses emotions of fear and anxiety in the subject that make the wish of her fidelity the anxious focus of the agent's desire-set. In other words, the desire that P is "turned on" emotionally by the threat and the suspicion about P not being true. At this juncture *the second* element becomes paramount: the subject is placed in circumstances such that acting in order to produce the state of affairs where P is made true is perceived as beyond her or his control. This is a typical circumstance for SD to start. When the subject is powerless about bringing about the desired state of affairs by a plan of action, he lingers in his thought. The fact that the fulfilment of his desire is beyond his reach, has the

effect of sinking the costs of inaccuracy in acquiring and processing data relevant to the truth of P. A properly diagnostic attitude will not mend the fact of Odette's infidelity for Swann.

4. Costs of Inaccuracy and Selective Vigilance

This point is nicely illustrated by Robert Jervis (Jervis 1976), who argues that the level of accuracy in gathering and processing information correlates with real-life incentives and prices for one's welfare. The costs of inaccuracy are irrelevant for Swann, for no attitude, vigilant or otherwise, will change Odette's conduct towards him. In similar fashion, Mele resorts to the notion idea of a focal error to be avoided (Friedrich 1993, Trobe and Liberman 1996). In case, no matter what the subject believes or does, he cannot bring about what he most wishes, the focal error to be avoided is not a dangerous outcome, which is rational to expect in case of incorrect information, but rather is exposing oneself to a painful truth *unnecessarily*, for changing the state of the world is beyond one's reach. Compared to Mele's explanation, Jervis adds a further non-cognitive component to the understanding of selective vigilance and the lack of it, namely the level of anxiety and stress concerning evidence processing. Low and high level of anxiety typically induce less accuracy and vigilance than a medium level of stress. But while low anxiety leads the subject to rely on routine and traditional patterns of conduct, high anxiety and stress tend to engender "defensive avoidance," that is a blocking of the negative information and reliance on a false soothing belief, which is precisely a form of SD. Clearly, this explanation holds only under the condition that the subject is an agent, though confused, and not a victim of the emotional state that triggers the biasing behind his back. Only an agent can sense, perceive, or suspect the need for vigilance.

To sum up: Odette's love and fidelity is beyond Swann's control. Hence the cost for inaccuracy sinks, since Swann has no expectation of changing Odette's mind through more accurate information. Besides, a positive correlation between high anxiety and defensive avoidance has also been shown: therefore, Swann not only cannot gain anything from a more accurate information of Odette's bearings, but his emotional state also tends to cause a cognitive attitude of "defensive avoidance", for, under his unfortunate circumstances, the focal error to be avoided is disbelieving Odette faithfully unnecessarily. Consequently, the threshold of evidence Swann requires to revise his belief about Odette's fidelity is disproportionally high. It is not a coincidence that typical examples of SD have to do with questions which are at the same time *crucial* for the subject's *well-being* and *beyond his control*: marital infidelity, fatal illness, and children's problems of addiction or criminal behavior. The first circumstance of crucial desires heightens anxiety, while the second makes the costs of accuracy sink.

These same two circumstances occur in twisted SD as well: a state of affairs crucial for the subject's well-being is threatened by the suspicion of contrary evidence that is beyond the subject's control. But the same circumstances may engender different forms of bias depending on different emotional and cognitive responses.

5. Revising the motivational state

Coming back now to the motivational state, I shall propose the following unpacking:

1) The operative desire is the desire that P and not the desire to believe that P.
2) The operative desire is a top priority in S's preference ranking.
3) The costs of inaccuracy in acquiring and processing data relevant to the knowledge that P is low or irrelevant for S.
4) The desire that P is emotionally loaded by fear and anxiety.
5) These two emotions are aroused by the appraisal or misappraisal of negative evidence threatening the truth of P.

The three issues left unsolved in the causal account, namely specificity, selectivity, and the unitary explanation of both straight and twisted SD, can now find a satisfactory response.

Briefly, (5) and (4) within the invisible hand model explain the specificity of SD. The desire that P is emotionally loaded by fear and anxiety produced by the appraisal of negative evidence as to the truth that P. S, anxiously suspecting that P is under threat, and powerless to counter the threat, starts thinking over the matter, and becomes prey to cognitive biases under the influence of her emotionally loaded wish that P. SD is entered only under condition (3), concerning the sinking of accuracy costs in the circumstances, and condition (4) inducing defensive avoidance. Thus, (3) and (4) jointly explain the selectivity of SD, but only under the condition that SD is intentionally entered. If by contrast, the subject is the victim of the motivational state getting a causal grip on cognition, activating biases, and producing the motivationally false P, then he can hardly be sensitive to costs of accuracy/inaccuracy and threshold of vigilance. Moreover, the biases' activation is secured by the motivational state ex ante, hence the costs of accuracy/inaccuracy seem redundant for starting SD.
Let's turn now to the unitary explanation of twisted and straight SD. The supporters of the deflationary account say that in either SD type, a motivational state triggers cognitive biases ending up in the fictional belief that P, either favorable or adverse. (Scott-Kakures 2000: 354). In twisted cases it is emotions rather than desires that switch on the biasing, and that explains the different result. Mele adds to this explanation a switch in the operative desire that is no more the desire that P (wife faithful) but the desire that Q (not being fooled). Thus this unitary model implies two unjustified shifts in twisted cases compared to straight ones: a) the emotional component takes the lead and causes the bias ending up in the false adverse belief; b) there is a shift in the operative desire as well, from the desire that P to the desire that Q. Both shifts are unexplained, and taken together they explain too much. If there is a shift in the operative desire from P to Q, then there is no need to suppose that the causal job is done by emotions, for the desire that Q is acting precisely like the desire that P in straight cases. Yet the resulting adverse belief ~P is not matching the operative desire that Q, thus giving rise to the issue of content-unrestrictiveness. If instead, the causal job is done by emotions, there is no need to shift the operative desire. It is still unexplained why S (wrongly) suspecting that P is under threat, becomes prey to the biases unleashed by fear and produces the false adverse belief ~P, and S[1](rightly) suspecting that P is under threat, let the desire that P to switch on biases producing the false favorable belief that P. The difference cannot lie in the fact that suspicion is right in straight cases and wrong in twisted ones, for it being right or wrong is an external condition and is apparently independent from the subject's motivational and epistemic setting.

6. Alternative explanation for both straight and twisted SD

As argued above, the motivational state has the same content for either case of SD. That S is moved by ungrounded emotions does not change how the SD process is started. In both cases, S is not (or believes not to be) in a position to act in order to undo the threat and bring about the state of affairs ex-ante. Like in most SD cases, S faces matters crucial to her well-being, yet beyond her control, actually lowering the costs of inaccuracy. The response is a thinking process affected by biases ending up in the deceptive belief. This is the common model operating in both cases; let us now consider the difference between straight and twisted SD. In straight SD, self-deceivers usually display confirmatory bias that is the mental habit to look for confirmation of one's hypothesis, while epistemic rationality would recommend falsification as the sound procedure. Confirmatory bias leads self-deceivers to ignore or discount the negative evidence and to focus instead on the positive data, ending up with the false but favorable belief that P. In twisted cases, instead, S imagines what she most fears in order to face the worst possible case. In principle, such a strategy is not faulty: it is justified by the insurance logic that is a strongly risk-averse approach of the kind: "let's assume the worst, and work out a maximin response in terms of attitudes, beliefs and actions to such a

scenario". The problem with this strategy, however, is that worst case scenarios become easily affected by a special bias—probability neglect that makes the reasoning faulty. In principle, worst-case scenarios are improbable events, representing only one extreme of a whole range of possibilities, yet it seems that once the scenario has been conjured up, the low probability is discounted, and the subject's fear and anxiety are reassured only by conclusive evidence showing the scenario false. Thus while S lowers the threshold of evidence to go on believing that P, S[1] heightens the threshold of evidence required to disbelieve ~P .The operative desire that P is matched by the resulting self-deceptive belief that ~P, though negatively, so as to be thematic, thus satisfying a commonly acknowledged requirement for SD.

The general explanatory model is the same in both cases: an emotional arousal following the appraisal, or the misappraisal, of negative evidence makes the wish that P the anxious focus of the subject; the active response to bring about P is foreclosed, hence the subject starts to think instead. The reasoning, influenced by the wish, attempts either at confirming the belief that P despite contrary evidence, or at disconfirming ~P, in the absence of the relevant evidence. In the first case, the reasoning is affected by confirmatory bias, and in the latter by probability neglect. In neither case the outcome is intended though produced by intentional thinking.

As much as the confirming strategy is the common mode of the lay hypothesis testing and is generally pragmatically reasonable (though epistemologically imperfect) similarly, the counter-factual mode of reasoning implied by the worst-case-scenario is, as said, at the basis of the insurance logic. In both types, normal pragmatic modes of reasoning are distorted by the anxious fear concerning P. As a result, in straight SD the false belief that P is confirmed, and in twisted SD, the opposite, but likewise false, belief ~P is produced. Admittedly, twisted SD does not serve the practical goal of dispelling S's worries, but it may perversely relieve a certain kind of anxiety, still fulfilling a purpose for especially anxious subjects. For such subjects usually display a preference for certainty over uncertainty even of unwelcome truths; consequently, they may be miserable, but will also feel a certain relief from the anxiety of suspecting the worst, and a (deceptive) sense of control.

7. Conclusions

Twisted SD cases pose a challenge for the intentional view of SD, according to which the self-deceptive belief fulfills a goal of the subject. Purposefulness cannot be assumed in the case of twisted SD, given that the self-deceptive belief opposes the subject's desires. It may actually be the case that extremely anxious subjects have a preference for unpleasant certainty over uncertainty. Yet, within the intentional framework it remains unclear why such overanxious subjects would not try to relieve their anxiety by believing something soothing, thus overcoming uncertainty and unwelcoming evidence at the same time. Causal accounts apparently present a clear advantage with reference to the issue of twisted cases. The latter need not be assumed as the outcome of special perversity. Twisted cases are a type of SD, produced in the same way as straight SD: they are the result of a motivational state causally switching on cognitive biases resulting in a false belief, no matter if favorable or adverse.

The appeal of such a unitary and simple account of both types of SD is clear. However, if this account is duly analyzed, then it turns out to be less unitary than claimed, for the causal switch is not the same in either case, and moreover, the operative desire in the twisted case is content-unrestricted.

My alternative invisible hand account provides a general model for either straight or twisted cases, which can be articulated in different reasoning strategies thus making sense of the two opposing outcomes as well. The invisible hand account holds that the SD process is entered intentionally and legitimately too: it is the response of an agent who has met with negative evidence concerning one of her desires, and is powerless to counter the threat. Under the circumstances, the agent is anxious and cannot act rationally, resolving her worries by fixing

the problem. The agent lingers in thinking about the problem, and thinking is something that she does intentionally. Yet the fact of her powerlessness, along with her anxiety, lowers costs of accuracy, manipulating the threshold of evidence required to either disbelieve or believe that P is true. The result is the relaxation of her usual epistemic standards without her noticing. This is the common track of SD of either type, made up of intentional steps and non-intentional twists; what changes is the cognitive strategy. If the subject is testing the truth of P, and in so doing, is caught in by confirmatory bias, we have straight SD, with the deceptive belief matching the agent's desire. If the subject conjures up a worst-case scenario, she falls victim to probability neglect, and without a final disproof she will be caught by the false and adverse belief that ~P. The choice of the strategy is intentional, but the related bias is not, just as the outcome is unintended.

To conclude: the same invisible hand model explains the self-deceptive production of both favorable and adverse beliefs, without the need to shift the causal components from desires to emotions, and without the need to shift the operative desire. The operative desire is always a) emotionally overloaded; b) the same that originates the worries of the subject; c) content-restricted in either case, though in twisted case it is in the negative form. If the product of twisted SD is an adverse belief, feeding worries and fears of the subject, it can still be said that in a special sense the belief serves a goal of the subject by dispelling her uncertainty over what to believe.

REFERENCES

Bermudez, J. L. (2000), "Self-Deception, Intention and Contradictory Beliefs" *Analysis*, 60, pp. 309-319;

Davidson, D. (1985), "Deception and Division", In LePore, E., and McLaughlin, B. (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Oxford, pp. 138-48;

Friedrich, J. (1993), "Primary Error Detection and Minimization (PEDMIN) Strategies and Social Cognition. A reinterpretation of Confirmation Bias Phenomenon", *Psychological Review,* 100, pp. 298-319;

Galeotti, A. E. (2012), "Self-Deception: Intentional Plan or Mental Event?" *Humana Mente*, 20, pp. 41-66;

Hayek, F. A. (1952), *The Counter-Revolution of Science*, The Free Press of Glencoe, Collier-Macmillan Limited, London;

Jervis, R. (1976), *Perception and Misperception in International Politics*, Princeton University Press, Princeton;

Lazar, A. (1997), "Self-Deception and the Desire to Believe", *The Behavioral and Brain Sciences*, 20, pp. 119-20;

Lazar, A. (1999), "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence'", *Mind*, 108, pp. 265-290;

Lynch K. (2013), "Self-Deception and Stubborn-Beliefs" *Erkenntis*, 78, pp. 1337-45;

Mele, A. (1999), "Twisted Self Deception", *Philosophical Psychology*, 12, pp. 117-137;

Mele, A. (2002), *Self-Deception Unmasked*, Princeton University Press, Princeton;

Neklin, D. (2002), "Self-Deception, Motivation and the Desire to Believe", *Pacific Philosophical Quaterly,* 83, pp. 384-406;

Nelkin, D. (2012), "Responsibility and Self-Deception: A Framework", *Humana Mente*, 20, pp. 117-39;

Nozick, R. (1977), "On Austrian Methodology", *Synthese*, 36, pp. 353-392;

Pears, D. (1984), *Motivated Irrationality*, Oxford University Press, Oxford;

Scott-Kakures, D. (1996), "Self-Deception and Internal Irrationality", *Philosophy and Phenomenological Research*, 41, pp. 31-56;

Scott-Kakures, D. (2000), "Motivated Believing: Wishful and Unwelcome", *Nous*, 34, pp. 348-75;
Scott-Kakures, D. (2002), "At Permanent Risk: Reasoning and Self-Knowledge in Self-Deception", *Philosophy and Phenomenological Research*, 65, pp. 576-609;
Talbott, W. J. (1995), "Intentional SD in a Single, Coherent Self", *Philosophy and Phenomenological Research*, 55, pp. 27-74;
Trobe, Y., and Liberman C. A. (1996), "Social Hypothesis Testing. Cognitive and Motivational Mechanism" In Higgins, E., and Kruglansky, A. (eds.), *Social Psychology: Handbook of Basic Principles*, Guildof Press, New York, pp. 239–270.

EDWARD HARCOURT
Oxford University – Keble College
edward.harcourt@philosophy.ox.ac.uk

MORAL EMOTION, AUTONOMY AND THE 'EXTENDED MIND'

abstract

This paper interweaves a 'micro' theme concerning shame and guilt and a 'macro' theme concerning self-regulation generally. Neither shame nor guilt is more other-independent than the other. Moreover, because other-dependence in either emotion is not a mark of heteronomy, neither emotion is more characteristic of a well-functioning moral consciousness. Then, relying on phenomena described by 'extended mind' theorists, I argue that a common view of self-regulation in children – that it is importantly other-dependent – is also true of adult self-regulation. But that is all the more reason to think that other-dependence and a well-functioning moral consciousness can go together. Moreover, since shame and guilt are one aspect of self-regulation, if other-dependence can be a characteristic generally of our well-functioning self-regulation – the 'macro' thesis – this supports the 'micro' thesis that other-dependence can characterize the well-functioning of both shame and guilt. The conclusion is that heteronomy lies not in the fact of other-dependence but in the nature of the dependence.

keywords

guilt, shame, self-regulation, autonomy, heteronomy, social cognition

Phenomenology and Mind, n. 11 - 2016, pp. 100-112
DOI: 10.13128/Phe_Mi-20110
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

The paper that follows tries to interweave two themes: a 'micro' theme, which concerns shame and guilt, and a 'macro' theme, which concerns self-regulation generally. I argue first of all that a common contrast between shame and guilt is frequently overdrawn, and that in fact neither emotion is more 'internal' or more other-independent than the other. Moreover, because other-*dependence* in either emotion is not *per se* a mark of heteronomy, neither emotion is *per se* more characteristic of a well-functioning moral consciousness than the other. I then turn to the 'macro' theme. Relying on phenomena that have been brought into the philosophical spotlight by advocates of the view that the mind is 'extended', I argue that a commonly held picture of the self-regulatory mechanisms of young children – namely that they are importantly other-dependent – is, plausibly, also true of the self-regulatory mechanisms of adulthood, though spelling this out in detail would need further research. If that is so, then – assuming only that the moral consciousness of typical adults can be well-functioning – there is all the more reason to think that other-dependence and a well-functioning moral consciousness can go together. Moreover, since shame and guilt are one aspect of the functioning of our self-regulatory mechanisms, if other-dependence can be a characteristic generally of our self-regulatory mechanisms when they function well – the paper's 'macro' thesis - then there is all the more reason to accept the paper's 'micro' thesis, that other-dependence can characterize the well-functioning of both shame and guilt. The conclusion to draw from both 'micro' and 'macro' theses is that the mark of heteronomy, and therefore the difference between the well- and the ill-functioning moral consciousness, is to be found not in the mere fact of other-dependence but in the nature of the dependence.

1. Rather like those strangely inaccessible renaissance paintings with titles like 'the contest between love and time', philosophical literature on guilt and shame has sometimes taken the form of a contest between the two emotions, asking which is the sign of a more developed or mature moral consciousness. Oddly, though opinions differ as to which of the two emotions is the mark of greater maturity, there is substantial agreement as to the standard by which maturity should be judged. The standard is *internality*, or *other-independence* or (inversely) *sociality*, the idea being that the more internal or other-independent, or the less social, an emotion, the less *heteronomous* it is. And so because – it is assumed – *autonomy* is the hallmark of moral maturity, the more internal or other-independent, or the less social, of the two emotions must be the more mature; or, more complicatedly, to the extent that the dominant or typical emotion in a moral consciousness is the less social or the more internal or other-

independent of the two, to that extent the moral consciousness is the more mature. I shall argue that this 'contest' not only has no winner but also is misconceived. I will stay with the terms of the 'contest', however, except in so far as for 'mature' I shall substitute 'well-functioning'. If 'mature' means 'characteristic of adults', the title bestows no honour because an adult moral consciousness can work otherwise than it should; if on the other hand 'mature' implies 'functions as it should', then we may as well say that instead.

The 'contest' I have in mind is nicely exemplified by AWH Adkins and Richard Wollheim, for though they hold contrasting positions on it, each position is structured by the same underlying assumptions. Williams summarizes Adkins's view of shame and guilt when he attacks it in *Shame and Necessity*:

> In the scheme of Kantian oppositions, shame is on the bad side of all the lines. This is well brought out in its ... association with the notion of losing or saving face. "Face" stands for appearance against reality and the outer versus the inner, so its values are superficial; I lose face or save it only in the eyes of others, so the values are heteronomous (Williams 1993: 77-78).

By contrast guilt, on the old-fashioned view, is closely associated with moral obligation, and with responsibility for one's own intentional actions. Thus, on this version of the contest, shame loses: it is less internal, and so more heteronomous, and so less mature than guilt (Adkins 1960). In *The Thread of Life*, Wollheim (1984) goes along with the privileged position occupied, on the old-fashioned view, by autonomy, and with the connection between autonomy and internality: a consciousness regulated by notions that are 'internal', or truly one's own, is morally more mature than a heteronomous consciousness, one regulated by the notions of others. However, shame (he argues) 'must be conceived to be just as much an interiorised sentiment as guilt is' (Wollheim 1984: 220). Indeed precisely because of guilt's association with moral obligation and the constraints it imposes on action, it is guilt not shame that is relatively heteronomous: the moral consciousness which is regulated by notions that are truly one's own is one whose dominant moral emotion is shame. In Wollheim's view, it is shame that is 'the prime moral sentiment of evolved morality' (Wollheim 1984: 220).

The differences between shame and guilt which supposedly underwrite a normative distinction between them – internal versus social, other-independent versus other-dependent and so on – are, however, insubstantial.

Wollheim's claim that guilt is less internal than shame is based on his distinction between two 'grades of internalization':

> There are figures that are merely internalized, and there are figures that are internalized and with which the person identifies (Wollheim 1984: 218).

That is, though Wollheim concedes – *pace* some critics of shame – that the basic audiences for both guilt and shame are internal, they are nonetheless different: whereas the audience for shame is one with which the subject is identified, the audience for guilt is *merely* internal. But what do these two grades of internalization amount to[1]?

Consider a case which is aptly described as an experience of 'merely' internal guilt. Albertine

1 In this and the following paragraphs, I adapt some material from my "Guilt, Shame and the "Psychology of Love"", In Louise Braddock and Michael Lacewing (eds.), *The Academic Face of Psychoanalysis*, Routledge, London, 2007, pp. 133-147.

always does housework on Sunday mornings. This is not because she likes housework – what she would really like to do on Sunday mornings is to lie in bed reading the newspaper. But Albertine was brought up in a strictly observant religious household. So although in adulthood she has not been a believer and never goes to church, she has found on Sunday mornings that if she tries staying in bed with the newspaper, guilt kills any pleasure she might otherwise have taken in it. Housework on the other hand – because it is burdensome – makes the guilt go away. Now there is a clear sense in which Albertine's guilt is 'merely internal'. It is *internal* in that it's not connected, for example, to whether *anyone else* is aware of what she does on Sunday mornings. But it is *merely* internal in that there is nothing in Albertine's current beliefs or values that *rationalizes* Sunday morning guilt even though, however much she rehearses the reasons why there's nothing wrong with Sunday morning relaxation, the habit of feeling doesn't shift. It is internal only topographically: it shapes Albertine's deliberation in the way an obstacle does, but the obstacle in this case happens to be inside her mind rather than, say, in her bloodstream (like a drug which induces nausea when she drinks alcohol)[2].

But though Albertine's guilt-avoidance strategy supplies a clear model for what 'merely internal' *might* mean in relation to guilt, it can't be what Wollheim means by this phrase. It's true that the beliefs and values which would rationalize Albertine's guilt are no longer ones with which she is identified, and that for this reason Albertine's strategy is heteronomous: her behaviour is a response to values that are not hers but someone else's (her parents', perhaps). But Albertine's story only makes sense if a certain history is in place, that is, *this* kind of failure to identify is only conceivable as a step down from some prior and now withdrawn identification. This does not fit the profile Wollheim describes for 'merely internal' guilt.

> [T]he development of, or development beyond, the superego [and a moral consciousness regulated by guilt] is best understood in this way: that the internal figure, or the group of internal figures, whose phantasised activities regulate the thoughts, feelings, and conduct of the person start off life as merely internalized figures - they 'confront' the ego - but, gradually ... come to be figures with whom the person is able to identify (Wollheim 1984: 220).

Wollheim, that is, envisages the 'merely internal' grade of internalization as developmentally prior to identification, whereas Albertine's case represents the afterlife of decayed identifications.

Let us turn, then, to Wollheim's own account of the forerunners of identification, the internal figures who – the phrase is Freud's - 'confront' the ego:

> [T]he dominant response of the infant to the superego is terror: the superego controls the child by means of fear. In the phantasies of internalization in which the superego occurs, it appears as a figure set over and against the child (Wollheim 1984: 201).

The internal figure sketched here differs from my picture of Albertine's 'merely topographical' internalization, but it too fails to match Wollheim's theoretical description of a lower grade of internalization, though for different reasons. A 'confronting' superego certainly goes with heteronomy: one can be motivated by terror without sharing the terrorizer's values. But

2 For more on this, see my "Mill's "Sanctions", Internalization and the Self", *European Journal of Philosophy,* Vol. 6, No. 3, 1998, pp. 318-334, and Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, Princeton, 2001.

it is unclear how fear turns into guilt just because the source of fear comes to be inside the subject's mind rather than outside it[3]; that is, the confronting superego seems *so* primitive that it leaves untold the story of how guilt makes its first appearance in consciousness. To see this, consider the citizens of a state which relies on terror as a means of social control. To begin with they try to predict what will attract punishment, and avoid doing it. In this phase they are motivated by fear, but the source of fear - the apparatus of state terror - is external. But their consciousness gradually undergoes a change: desires, plans, initiatives which previously they formed but refrained from carrying out now no longer even occur to them. This change represents the internalization of terrorizing external figures: the citizens' behaviour is still regulated by fear, but the regulation no longer requires actual threats by secret policemen and the rest. What matters here is that there's a word for what results when a terrorizing other is moved inside the mind, and it's not guilt-proneness but merely *timidity*; that is, the citizens' self-censorship makes perfect sense without the supposition that the plans they no longer make are objects of guilt. Internalizing the terrorizing figures makes *a* difference, namely the difference between a disposition to avoid punishment that is geared to actual threats and a generalized disposition to avoid it that indeed is no longer thought of as punishment-avoiding but simply as the way to behave. But that isn't the right *sort* of difference to explain how guilt appears on the scene.

In summary, Wollheim tries to distinguish two grades of internalization such that guilt is already present at the lower of the two grades while shame appears only at the higher one. But though Albertine's case includes guilt and is distinct from Wollheim's higher grade in that identification is absent, it cannot represent Wollheim's lower grade because it is not developmentally prior to identification. The police state case, on the other hand, shows that a consciousness may be regulated by internalized fear without being regulated by guilt. Moreover when we imagine guilt appearing on the scene, what we seem to be imagining is identification with the agents of terror. So we jump from a lower grade of internalization at which guilt is absent to a higher grade at which it's present, but where identification is present too. Shame and guilt therefore do not occupy two distinct grades of internalization, so that is not a difference that can underwrite a normative distinction between them.

2.

I want to turn now to two further alleged differences between guilt and shame, each of which plays its part in the 'contest' between them, and which in different ways concern their relation to heteronomy. The first concerns the allegedly tight connection – whether deployed in praise of guilt (as on the old-fashioned view Williams criticizes) or against it (as in Wollheim) – between guilt and morality. The second concerns the relations of the two emotions to other people.

One may feel shame equally properly at cheating in an exam, raising a laugh by telling a bad joke to an undiscriminating audience – both of which are intentional, but only the first of which I take it involves any moral fault - falling over on stage, being overweight (unintentional) or having a prominent scar (involuntary). So shame is only loosely connected with the intentional or indeed the voluntary, or with morality. But is the connection any closer for guilt? No: proper objects of guilt include breaking rules of religious observance, bringing an insistent non-member into the 'members only' room at a club, failures of prudence such as not visiting the dentist, and breaches of self-imposed rules such as diets or work schedules, none of which are moral wrongs. People also feel guilty about their own good

3 The same point is made by David Velleman, "A Rational Superego", in his *Self to Self*, Cambridge University Press, Cambridge, 2006, p. 135.

fortune. One's enjoyment of a comfortable house to live in can be spoiled by the distinctively guilty thought that there are other people who are homeless. But the guilty thought does not seem to need the mediation of any further thought as to a route by which, were one to live less comfortably, homelessness might be mitigated. Indeed, far from having to be 'I have deprived them', the guilt-inducing thought can be 'what have I done to deserve this?'. In other words, the guilt-inducing thought can focus precisely on the *absence* of any doing of one's own that warrants the difference between one's own advantage and others' disadvantage. So the link between guilt and the intentional or indeed the voluntary – one might unexpectedly inherit a lot of money - is also looser than is often supposed. This is not of course to recommend guilt in such cases, but just to say that one's own unearned advantage is an *intelligible* object of guilt; a parallel point could be made, say, about shame at having a scar. So on assumptions that structure the 'contest' – the closer the connection to morality or to the intentional, the better-functioning the moral consciousness (Adkins) – once again neither guilt nor shame emerges as a clear winner[4].

As to other people, let me rehearse briefly the reasons why the fact that shame, though admittedly other-dependent is not - *pace* the old-fashioned view - always heteronomous. To be sure, shame is often occasioned by the reaction, or the imagined reaction, of others. But this observation doesn't support the view that shame is *per se* a heteronomous emotion. To see this, we need only ask why other's opinions of our actions should matter to us – in particular, why they should matter to us in the special way such that imagining whether others would try to shame (or praise) us for doing such-and-such informs us *directly* whether or not our doing such-and-such would be shame*ful* (or praise*worthy*) – except on the supposition that we and they, at least as far as this action is concerned, have something relevantly in common. This community of values with others is what shame*less* people precisely lack, which is why attempted shaming by others makes no difference to them – at least not the relevant difference. Of course the ill opinion of others may matter even to the shameless, just as being hit by a stone while in the stocks hurts whether or not one cares why it was thrown. But neither the stone nor others' ill opinion will hurt in the distinctive way the stone hurts a victim whose values are in tune with the stone-thrower's, namely as an expression of what both of them agree is the shamefulness of whatever it is that explains the victim's being put in the stocks. Just the same point goes, *mutatis mutandis*, for the others' good opinion. Although ordinary speech can be very free with the phrase 'approval seeking', others' approval functions motivationally in very different ways depending on how well adjusted to one another are the values of the seeker and the person from whom approval is sought. At one extreme, only a shameless person could enjoy being applauded for a prize-winning work they had plagiarized, and what they would enjoy in that situation would be (say) being the object of attention of a room full of people, being photographed etc., *not* applause as the expression of others' recognition of the value they themselves see in what they have done (because they haven't done anything). Moreover, for the plagiarist, the good of being applauded is a *self-standing* reason for passing off the plagiarized work as their own – it's only good as a means to securing the applause. At the contrasting extreme, applause is not a self-standing reason: the agent's reasons for seeking it are the very same as the audience's reasons for applauding,

4 Indeed one might argue that the very intimacy between guilt and the 'morality *system*' (Bernard Williams, *Ethics and the Limits of Philosophy*, Fontana, London, 1985, *passim*), with its meritocratic emphasis on desert, is what explains the unexpected looseness of the connection between guilt and the intentional (or indeed the voluntary): the morality system does not so much restrict guilt to one's intentional wrongs – which would be deserved – as foster a generalized horror of the undeserved. So confining rewards to the deserved and feeling guilt at the undeserved are two sides of the same coin.

namely their shared belief in the value of the work. Somewhere short of this extreme belong a class of cases which, in a way, make the same point even better. These are cases where there is a community of values between the agent and his or her audience, but where the audience makes the difference to the agent's motivation to act on them. Thus George Eliot describes Esther - heroine of *Felix Holt* – as viewing her beloved Felix as 'another better self' (Eliot 1972: 591). Far from acting contrary to her own judgment under pressure of the desire for her lover's approval – which of course is possible too - Esther's knowledge of Felix's approval motivates her in the way it does because he speaks, more loudly than she does herself, for what she values.

Now just where we might expect to find a contrast between shame and guilt, there's in fact an exact analogy. Guilt, like shame, can be absent until it is brought on by the judgment of others – a court verdict, for example. Nonetheless, when the judgment of others does work like this, it is not their judgment alone but that judgment *in the context of something the subject shares with those others* that makes it work. Absent that shared something and the others' judgments of wrongness won't work, or won't work in the relevant way. Someone who feels alienated from the entire judicial system won't feel guilt when denounced by a judge, though they may feel exasperation, fear and a range of other things, emotions which stand to guilt in much the way as the pain cause by the flying stone stood to shame in the earlier illustration. Again, the all-seeing eye of the superego (Freud 1985: 317) is sometimes said both to be central to the phenomenology of guilt *and* to show why guilt is less social an emotion than shame. In fact it seems to show the exact opposite, by emphasizing the centrality to the phenomenology of guilt of *being known* to have performed some act by another who shares with one the beliefs that suit it to be the object of the emotion[5]. The same phenomenology is at work in reverse when we see people staving off the experience of guilt by concealing what they have done.

My first conclusion, then, is that shame is neither less internal or other-independent, or more social, than guilt; nor, alternatively, is it less social, more internal or other-independent: both are importantly dependent on the opinions or judgment of others.

Now this would be a comparatively uninteresting conclusion if all it showed was that the familiar 'contest' between guilt and shame has no clear winner. But it shows more than this. The point of the 'contest' was not merely to compare guilt and shame along these various dimensions, but to compare them along these dimensions because of a presumed connection between internality or other-independence and - the 'winning' characteristic – the well-functioning moral consciousness.

The cases I have discussed, however, show that the link between internality or other-independence and the well-functioning moral consciousness doesn't hold. Guilt can be internal and yet part of a moral consciousness that doesn't function well, as with Albertine. On the other hand where the opinion or judgment of others matters in the distinctive shame- or guilt-evoking way – that is, where the others share with the subject the values in the light of which some feature of the subject, or of one of his or her actions, is the object of the emotion - both guilt and shame are social not internal, other-dependent not other-independent, and yet the subject's moral consciousness is working exactly as it should[6].

5 As with the citizens of the police state again: if the seen did not have something relevantly in common with the all-seeing, then why would being seen *be* part of the phenomenology of guilt, rather than of the phenomenology of fear?

6 This is perhaps the point at which to notice that the standard for a well-functioning moral consciousness is not itself a moral standard. The shameless plagiarist's moral consciousness is functioning badly with respect to (say) the value of literary achievement. But if in her view failure to be glamorous is so shameful that it makes plagiarism worthwhile, and if there are also audience members who applaud her glamour rather than her (apparent) achievement, then with respect to *that* value, her moral consciousness is functioning well. But that is not to say she is

What is more, these cases show that the distinction between other-dependence (sociality) and other-independence (internality) aligns poorly with the distinction between autonomy and heteronomy: Albertine's self-regulatory mechanisms are internal yet she is surely heteronomous, because her disposition to feel guilt is not responsive to her own values. On the other hand though pride and shame in the case of George Eliot's Esther (or guilt, in the case of the criminal not alienated from the judicial system) are thoroughly social, because the values they respond to are those of their respective audiences, there is nothing heteronomous about these agents, because the values in question are not *just* those of others, but also their own. That there should be a connection between autonomy and a well-functioning moral consciousness, or heteronomy and an ill-functioning one, is highly familiar. That there should be compatibility between autonomy and sociality or other-dependence is less so. To try to understand why that is, and to suggest why things should be otherwise, I now turn to my 'macro' theme of self-regulation generally and – slowly - to the 'extended mind'.

3. First I want to present a narrative about self-regulation which crops up in places as ostensibly different as Kant and the psychoanalysis of children. It goes like this. The infant or child does not regulate itself – its behaviour, emotions and so on – but rather is regulated by a relation to another, or others. This regulation begins (very roughly) in mutually attuned interaction between infants and their first caregivers. But as maturation progresses, the role formerly played in the regulation of behaviour (etc.) by others gradually gets replaced by an internal mechanism in the operation of which others play no part. This kind of view has its *locus classicus* perhaps in Kant, where reliance on another ('a book which understands for me, a pastor who has a conscience for me, a physician who decides my diet'[7]) is the norm in childhood but written off as a moral failing in anyone who has passed beyond that stage[8]. But D.W. Winnicott too describes the mother's role in relation to the infant as that of an 'auxiliary ego', needed in order to perform essential ego-functions while the child's own ego is not yet fully formed[9]. The implication is that once the child's ego *is* fully formed, it does whatever it needs to do all on its own – no auxiliaries needed. But of course in Winnicott implicitly, and in Kant explicitly, the transition from regulation by others to self-regulation is the transition to autonomy. This narrative is thus a spectacular example of how the connection between autonomy and other-independence, which structures the 'contest' between shame and guilt, is made: the less the dependence on others, the greater the autonomy.

Now for the second piece of new material. In some interesting recent work, Joel Krueger has presented evidence to support the idea that the mind is extended (Krueger 2011, Varga and Krueger 2013). Krueger, like me, is interested in the phenomenon of self-regulation, so he is struck by the fact that children seemingly regulate themselves using various things beyond themselves, including music (as when music helps them to calm down and get to sleep), cuddly toys and, most impressively of all, other people. According to Krueger, if the right thing to think about Clark and Chalmers's famous diary (Clark and Chalmers 1998) is that the mind encompasses *that* (because memory is a mental function and memory depends on the diary), then, because the same can be said of self-regulation, mind also extends to teddy bears,

not making a moral mistake in valuing glamour so highly.

7 I. Kant, "What is Enlightenment?", in his *Critique of Practical Reason and Other Writings in Moral Philosophy*, tr. and ed. L. W. Beck, University of Chicago Press, Chicago, 1949, p. 286.

8 If 'nature has long since discharged [me] from external direction', the cause is 'not lack of reason but ... lack of resolution and courage to use it without direction from another', *ibid.*

9 D. W. Winnicott, 'ego immaturity is naturally balanced by ego-support from the mother' ("The Capacity to be Alone", *The Maturational Processes and the Facilitating Environment*, Karnac, London, 1990, p. 32).

lullabies and other people. Now here I am not interested in the extended mind idea *per se*, but only in the phenomena Krueger adduces in support of it: whether they do support the idea, and if so on what interpretation of it, is a subject for another day[10]. But I take it that these phenomena *do* show that self-regulation in children is a business that crucially involves the child's environment, including other people.

Two questions raised by Krueger's work that cannot be set on one side, however, are these: first, to what extent is the involvement of others in self-regulatory mechanisms always (in a sense I shall explain) instrumental? And secondly, to what extent do Krueger's claims about the distributed character of self-regulation carry over from children to adults?

With regard to instrumentality, Krueger describes music as '*a tool* ... to enact micro-practices of emotion-regulation', a '*crucial tool* for cultivating and regulating our emotional and social life' (my italics) (Krueger 2011: 2-3). Now it's not only music that can be a 'tool' of self-regulation: other human beings can be tools too, as when I control my intake of whisky by persuading a friend to hide the key to the drinks cabinet. But for the purposes of arguing that things or persons beyond the body are part of one's self-regulatory mechanisms, this kind of case is not very helpful. For the regulatory role of music when it's a tool, or of the other person, is mediated by a fully formed intention (to get to sleep, say, or to drink less) and the external thing comes in simply as a means to that. When I intend to cut down on whisky for the same of my health, however, I am already a self-regulator, but do not 'use' the reason of health (or the thought of my future health, or whatever) as a means to cutting down. So it looks as if, if external things as well as thoughts, reasons etc. are really to belong to the mechanisms of self-regulation, as opposed to being things which these mechanisms can make use of now and again, they must feature in self-regulation in a more than merely instrumental way. But this, indeed, is just what we find in the case of music and infants. When caregivers play music to infants to get them to sleep or to calm down, there is instrumentality on the caregivers' side, but not on the infant's: in listening to the music the infant is simply regulating itself, not listening *in order to* regulate itself. This brings me to my second question: does anything like this interesting case of the non-instrumental involvement of an external thing in self-regulation occur in adulthood too?

Krueger's answer is yes. 'Though synchrony in distributed emotion regulation is usually considered as a developmental phenomenon', he says, 'there is growing evidence that dyadic adult relationships have similar functions, providing a context for distributed emotion regulation throughout life' (Varga and Krueger 2013: 286). Nonetheless his positive answer is qualified: distributed emotion regulation 'also characterize[s] some adult relationships like intimate romantic relationships and the one between therapist and client' (*ibid.*). The qualification reflects a broader tentativeness in the literature. 'Throughout the life cycle, through the emotion-based interaction with a sensitive ... other, people are able to manage that which is felt to be too much to do alone', writes Diane Fosha (Fosha 2001: 229). But in the very next sentence, she says that 'eventually, [people] become able to do for themselves what was initially accomplished through the relationship' (*ibid.*). The fact that self-regulatory mechanisms are distributed in a relationship with a *therapist* suggests that when this happens in adulthood, it remedies a deficit, thus reaffirming not only the questionable idea that the course of successful psychotherapy recapitulates normal psychological maturation[11], but also

10 For various interpretations, see M. Wheeler, "Revolution, Reform, or Business as Usual? The Future Prospects for Embodied Cognition", In L. Shapiro (ed.), *The Routledge Handbook of Embodied Cognition*, Routledge, London, 2014, pp. 374-383.

11 For questions, see Stephen Mitchell, *Relational Concepts in Psychoanalysis*, Harvard University Press, Cambridge MA, 1988, and my "Madness, Badness and Immaturity: Some Conceptual Issues in Psychoanalysis and Psychotherapy", *Philosophy, Psychiatry and Psychology*, forthcoming.

of course the Kantian-psychoanalytic narrative rehearsed earlier according to which the well-functioning self-regulatory mechanisms of adulthood are a solitary affair[12].

Now clearly, in order to argue that other people are part not only of children's self-regulation systems, but part of adults' self-regulation systems too – quite standardly, and not just in the deficit context of psychotherapy, or even only in intimate relationships – evidence is needed, and I certainly can't provide much of that here, so to that extent the 'macro' thesis of this paper is dependent on future research. Let me offer, however, some reasons for thinking that the phenomenon Krueger identifies as true of immature self-regulation generally is also a widely generalized phenomenon in adulthood.

It seems to me that the phenomenon of involving others non-instrumentally in self-regulation is pervasive in adult life, and it does not do justice to this phenomenon to look for it only in special relationships[13]. At the risk of stating the obvious, *telling an intimate one's troubles* regulates one's trouble, both in the sense that it may help one to express it (and so to deal with it as one would not be able to if the troubles were unexpressed) and in the sense that, often, it makes one less troubled. But this truism is but a step away from another one that does not fit the 'intimate or therapist' model, for one can achieve the same effect my telling a non-intimate friend, a colleague, an acquaintance, a sympathetic stranger one gets into conversation with on a train – the relationships in question really do not have to be very special at all. Also don't forget that regulation can be up as well as down: my pleasure in a memory can be intensified by sharing it.

In fact there seem to be so many examples of the involvement of non-special others in adult self-regulation that it is hard to know where to begin. Group activities, where the others in question may be total strangers and united only by the fact that they too are taking part, have an important self-regulatory function. Shouting for their team at a football match, for example, makes people 'feel better'; group singing – compare Krueger on music for infants - whether it be part of a collective process of mourning or of mutual encouragement (as in marching songs) is a powerful emotion regulator. Another extremely ordinary example is that many people are less able to carry on ordinary activities effectively when alone. But the remedy for being alone need not be intimacy, but simply the presence of others going about their business. Echoing Krueger and Varga's notion of 'proximity'[14], a lecturer may speak more confidently if there are a few people in the audience she recognizes, even if she doesn't know their names, so the presence of known non-intimates regulates stress (O'Donovan and Hughes 2008). A similar psychology underlies the fact that non-intimates go along to lectures, court hearings or other 'difficult' encounters in order to 'offer their support'; people feel supported in difficult endeavours by *wearing* the same things as complete strangers – military uniform, for example, or badges and armbands on a protest march. Here I have touched only on phenomena involving the regulation of emotion, but sharing one's state of mind with other people enables us to do many things we couldn't do alone, and thereby regulates our thoughts and our behaviour[15]. None of these phenomena involve 'special' others; as they are such a

12 Its prevalence is also noted by L. Diamond and L. Aspinwall, "Emotion Regulation Across the Life-Span", *Motivation and Emotion*, Vol. 27, No. 2, 2003, pp. 125-156. See also L. Diamond and L. Aspinwall, "Integrating Diverse Developmental Perspectives in Emotion Regulation", *Motivation and Emotion,* Vol. 27, No. 1, 2003, pp. 1-6.

13 For material on both intimate and non-intimate relationships, see J. Coan, "The Social Regulation of Emotion", In J. Decety and J. Cacioppo (eds.), *The Oxford Handbook of Social Neuroscience*, Oxford University Press, Oxford, 2011, pp. 614-623.

14 'Our basic feeling of interpersonal connectedness', which 'allows individuals to morph into a cohesive dyadic system able to carry out *distributed emotion regulation*' (Krueger and Varga 2013: 272).

15 See P. Pettit, "Substantive Moral Theory", In E. F. Paul, F. Miller Jr., and J. Paul (eds.), *Objectivism, Subjectivism and Relativism in Ethics*, Cambridge University Press, Cambridge, 2008, pp. 1-27.

familiar part of adult life, there seems no reason to say they only remedy deficits either. (The desire to share one's experiences with others is not a deficit).

4.

Though of course more work is needed, the phenomena to which 'extended mind' theorists have helped to draw our attention point the way to the view that the interpenetration of self and other is - as the paper's 'macro' thesis has it - a pervasive feature of adult self-regulation, not just an exception to normal adult self-regulation constituted by intimates, or a prop supplied to ill-functioning adult self-regulators by psychotherapists. But now assuming that the capacity for self-regulation is part of autonomy, we have a conclusion that stands in opposition to the idea – that we find in both Kant and in the psychoanalytic tradition – that autonomy precludes dependence on others. Now that autonomy and other-dependence should not be a puzzling combination was, of course, the conclusion of the earlier discussion of guilt and shame. The convergence should be no surprise. Though there is a great deal more to self-regulation than just the disposition to feel these two emotions, shame and guilt are emotions of self-regulation[16]. By this I do *not* mean that they are disagreeable experiences the avoidance of which motivates us to be careful self-regulators: though that *can* happen, it is very much a deviant case[17]. What I mean is rather this. Of course if we were all perfect self-regulators – if we never acted (or thought or felt) otherwise than as we think we should – we would have far fewer occasions to experience either emotion than each of us typically does. (I do not say *no* occasions, thanks to the weakness of the link between either emotion and the voluntary – see above.) But for imperfect creatures like ourselves, guilt and shame – at least when they are experienced in relation to the right object, to the right degree, at the right time and so on – are signs that our self-regulatory mechanisms are in good working order. But now if the paper's 'macro' thesis is true, it should be no surprise that in the particular cases of shame and guilt, when both emotions are working as they should, they can – as the paper's 'micro' thesis has it – both be social or other-dependent and characteristic of a well-functioning moral consciousness.

The paper has only argued, of course, that autonomy and other-dependence can go together, not that they always do. That is why it is still open to us to say, in line with conventional wisdom, that notwithstanding the phenomena Krueger describes, small children are *not* autonomous self-regulators. Nonetheless if indeed they are not, this cannot be because of the mere presence in their self-regulation systems of other people, but rather (for example) because of the typical asymmetries in knowledge and power of adults and children, a point which suggests that the conventional wisdom is an exaggeration: it is not as if there is no such thing as an adult's having undue influence over a child - even a very small one - or interfering with its freedom[18]. Something very similar will go for adults: the interpenetration of self and other in adulthood seems to be neutral between autonomy and heteronomy, so when adults are autonomous, that will not be because their self-regulation systems *don't* involve others – not only, for example, because their dispositions to feel guilt and shame are not sensitive to values they share with them - but because they don't involve others in certain ways: if the others don't have undue influence, don't exploit emotional connections which subvert the other adult's reason, don't terrify or threaten them, and so on. A proper taxonomy of these

16 This term is greatly preferable to 'conscience', because of the moral associations of the latter together with the fact that the standards in the light of which we regulate ourselves, and such that guilt and shame express in part our failings in so doing, are so much more than moral – as witness the examples earlier in this paper.

17 An example is Albertine's case above.

18 For autonomy in infancy, see my "Attachment, Autonomy and the Evaluative Variety of Love", In E. E. Kroeker and K. Schaubroeck (eds.), *Love, Reason and Morality*, Routledge, London, 2017, pp. 39-56.

kinds of way cries out to be constructed[19]. Such a taxonomy would enable us to leave behind internality or other-independence as measures of the well-functioning moral consciousness, and point us towards a grown-up account of the contrast between autonomy and heteronomy.

REFERENCES

Adkins, A. W. H. (1960), *Merit and Responsibility: A Study in Greek Values*, Clarendon Press, Oxford;
Clark, A., and Chalmers, D. (1998), "The Extended Mind", *Analysis,* 58, pp. 7-19;
Coan, J. (2011), "The Social Regulation of Emotion", In Decety, J., and Cacioppo, J. (eds.), *The Oxford Handbook of Social Neuroscience*, Oxford University Press, Oxford, pp. 614-623;
Diamond, L., and Aspinwall, L. (2003), "Emotion Regulation Across the Life-Span", *Motivation and Emotion*, Vol. 27, No. 2, pp. 125-156;
Diamond, L., and Aspinwall, L. (2003), "Integrating Diverse Developmental Perspectives in Emotion Regulation", *Motivation and Emotion*, Vol. 27, No. 1, pp. 1-6;
Eliot, G. (1972), *Felix Holt*, Coveney, P. (ed.), Penguin, Harmondsworth;
Fosha, D. (2001), "The Dyadic Regulation of Affect", *Journal of Clinical Psychology*, Vol. 57, No. 2, pp. 227-242;
Freud, S. (1985) *Civilization and Its Discontents*, in *The Pelican Freud Library vol. 12: Civilization, Society and Religion*, Dickson, A. (ed.), Penguin, Harmondsworth;
Harcourt, E. (forthcoming), "Madness, Badness and Immaturity: Some Conceptual Issues in Psychoanalysis and Psychotherapy", *Philosophy, Psychiatry and Psychology*;
Harcourt, E. (2017), "Attachment, Autonomy and the Evaluative Variety of Love", In Kroeker, E. E., and Schaubroeck, K. (eds.), *Love, Reason and Morality*, Routledge, London, pp. 39-56;
Harcourt, E. (2007), "Guilt, Shame and the "Psychology of Love"", In Braddock, L., and Lacewing, M. (eds.), *The Academic Face of Psychoanalysis*, Routledge, London, pp. 133-147;
Harcourt, E. (1998), "Mill's "Sanctions", Internalization and the Self", *European Journal of Philosophy*, Vol. 6, No. 3, pp. 318-334;
Kant, I. (1949), "What is Enlightenment?", Id., *Critique of Practical Reason and Other Writings in Moral Philosophy*, Beck, L.W. (tr. and ed.), University of Chicago Press, Chicago;
Krueger, J. (2011), "Doing Things with Music", *Phenomenology and Cognitive Science*, Vol. 10, No. 1, pp. 1-22;
Mitchell, S. (1988), *Relational Concepts in Psychoanalysis*, Harvard University Press, Cambridge MA;
Moran, R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, Princeton;
O'Donovan, A., and Hughes, B. (2008), "Access to social support in life and in the laboratory", *Journal of Health Psychology*, 13, pp. 1147-1156;
Pettit, P. (2008), "Substantive Moral Theory", In Paul, E. F., Miller Jr., F., and Paul, J. (eds.), *Objectivism, Subjectivism and Relativism in Ethics*, Cambridge University Press, Cambridge, pp. 1-27;

19 Another taxonomy which cries out for construction is a taxonomy of the ways in which others can be involved in my self-regulatory mechanisms. Arguably they are involved somehow when I violently assault a stranger as a way of coping with my own unmanageable feelings of distress or (to take a less extreme example) when I arrive home in a bad mood, shout at my wife for (as I later see) something for which she is not to blame, upset her – and then cheer up. In psychoanalytic terms, I have 'projected' my bad feelings into her thereby ridding myself of them; in any case, she has 'helped' to up-regulate my mood, as has my victim in the assault example. These look psychologically very different from telling her why I am in a bad mood and cheering up that way (or from finding someone to tell about my unmanageable distress). A taxonomy of these seemingly different psychologies of other-dependence is also urgently needed, and in such a way as to make it clear how it maps on to the taxonomy of undue influence etc. just mentioned.

Varga, S., and Krueger, J. (2013), "Background Emotions, Proximity and Distributed Emotion Regulation", *Review of Philosophy and Psychology*, Vol. 4, No. 2, pp. 271-292;
Velleman, D. (2006), "A Rational Superego", Id., *Self to Self,* Cambridge University Press, Cambridge, pp. 129-155;
Wheeler, M. (2014), "Revolution, Reform, or Business as Usual? The Future Prospects for Embodied Cognition", In Shapiro, L. (ed.), *The Routledge Handbook of Embodied Cognition*, Routledge, London, pp. 374-383;
Williams, B. (1985), *Ethics and the Limits of Philosophy*, Fontana, London;
Williams, B. (1993), *Shame and Necessity*, University of California Press, Berkeley/Oxford;
Winnicott, D. W. (1990), "The Capacity to be Alone", Id., *The Maturational Processes and the Facilitating Environment*, Karnac, London, pp. 29-36;
Wollheim, R. (1984), *The Thread of Life*, Yale University Press, New Haven/London.

MARIA GIOVANNA BEVILACQUA
Università della Svizzera italiana
giovanna.bevilacqua@usi.ch

CARING ABOUT ETHICS OF CARE: A NEW DIMENSION

abstract

Can ethics of care paradigm by Carol Gilligan provide a helpful contribution to tackling some of the contemporary world social issues? We can attempt to answer this question by focusing on some themes that ethics of care underlines. We have to focus on the responsiveness of the moral subject for the others' needs and also we have to focus on the difference between concrete other and generalized other. Together with Carol Gilligan, Virginia Held, Seyla Benhabib and Charles Taylor indicate the way to face this philosophical argument.

keywords

care, responsiveness, feelings, other, recognition

Phenomenology and Mind, n. 11 - 2016, pp. 114-119
DOI: 10.13128/Phe_Mi-20111
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

A paradigm

In 1982 Carol Gilligan published the work *In a Different Voice: Psychological Theory and Women's Development*. Since then, this book has been considered the cornerstone of women's studies. In this book and in the empirical research it was based on, Carol Gilligan showed that women's moral orientation is different from men's. Lawrence Kohlberg's work on moral development, based on Jean Piaget's theory, distinguished three levels in people's moral development: *preconventional level*, *conventional level* and *postconventional level*. In Kohlberg's progression pattern the first level is the moral heteronomy, the highest level is the moral autonomy. Lawrence Kohlberg said that human beings reach the highest stage of moral development when (and if) they are able to judge and act in accordance with abstract and universal justice-equity oriented principles denying feelings and emotions as motivation to action. According to Lawrence Kohlberg's empirical research, based on interviews, many men reach the highest moral stage but very few women do. In Carol Gilligan's opinion the differences represented in psychological literature as steps in a developmental progression are two different moral orientations: a morality focused on principles, ethics of rights, and another kind of morality, ethics of care. Gilligan's empirical research based on interviews, resulted in two main outcomes:

a) women's moral point of view is always linked to the awareness that human beings have relationships with other human beings;

b) in order to solve a moral dilemma, besides reason, women take into account emotions, feelings and caring for people and for relationships while men take into account reason and justice as equity.

These facts cause women to be at lower levels in Kohlberg's justice oriented moral development model. Lawrence Kohlberg equated moral development with the development of rights reasoning. Gilligan's research and Kohlberg's interviews concerning solving moral dilemmas were about this moral dilemma: Mrs. Heinz is so dangerously ill than she may die. Mrs. Heinz needs a very expensive drug. Mr. Heinz isn't rich. The druggist refuses to lower drug's price. Should Mr. Heinz steal the drug for saving his wife's life? A teenager boy answered that Mr. Heinz should steal the drug; a teenager girl answered that Mr. Heinz shouldn't steal the drug. For the boy "Mr. Heinz should steal the drug because a human life is worth more than money and if the druggist only makes 1000 dollars, he is still going to live, but if Heinz doesn't steal the drug, his wife is going to die" (Gilligan 1982: 25). For the girl "there might be other ways besides stealing the drug, like if he could borrow the money or make a loan or something [...]. If he stole the drug, he might save his wife then, but if he did,

he might have to go to jail, and then his wife might get sicker again, and he couldn't get more of the drug, and it might not be good. So they should really just talk it out and find some other way to make the money". (Gilligan 1982: 26).
The boy answered according to justice as equity, universal and rational principles (Heinz should steal the drug because life right is more important than property right), the girl answered according to feelings, emotions and care for persons and relationships (Heinz shouldn't steal the drug because he might go to jail and he might break the love relationship with his wife, so the joy for the recovery of Mrs. Heinz could become sadness and sickness again). Women's ethics is an *ethic of care*. Ethics of care cannot be strictly considered as women's ethics but as an *ethic theme* (also a man can act according to ethic of care). What do we mean by ethics of care? Carol Gilligan defines ethics of care in these terms: "in this conception the moral problem arises from conflicting responsibilities rather than from competing rights and requires for its resolution a mode of thinking that is contextual and narrative rather than formal and abstract. This conception of morality as concerned with the activity of care centers moral development around the understanding of responsibility and relationships, just as the conception of morality as fairness ties moral development to the understanding of rights and rules" (Gilligan 1982: 19).
Together with Carol Gilligan, one of the most important theorists of ethics of care is Joan Tronto. Recently has defined ethics of care in these terms: "an ethic of care is an approach to personal, social, moral, and political life that starts from the reality that all human beings need and receive care and give care to others. The care relationships among humans are part of what mark us as human beings. We are always interdependent beings" (Tronto, Interview on August 4th, 2009). We can define this ethic theme as a paradigm by means of two considerations: morality means taking care of someone/taking care of the relationships one is involved in, and morality is not only based on abstract and rational principles. We can focus the paradigm on some points: 1) morality means responsiveness to others' needs. Care is responsiveness/care is responsibility; responsiveness/responsibility is to be able to satisfy someone's demand. Morality is founded "in a sense of concrete connection and direct response between persons, a direct sense of connection which exists prior to moral beliefs about what is right or wrong or which principles to accept" (Blum 1988: 477); 2) morality doesn't only come from reason, but also from sensibility/emotions/feelings. "Understanding the needs, interests, and welfare of another person, and understanding the relationship between oneself and that other requires a stance toward that person informed by care, love, empathy, compassion and emotional sensitivity" (Blum 1988: 475); "morality necessarily involves an intertwining of emotion, cognition, and action, not readily separable [...]. Caring action expresses emotion and understanding" (Blum 1988: 476); 3) the moral agent is always *embedded/encumbered*, the moral patient is always *embedded/encumbered*, too. Each human being is linked to some other human being, place and situation. Each human being is different from every other human being. The "moral self is radically situated and particularized. It is defined by its historical connections and relationships. The moral agent does not attempt to abstract from [...] particularized self to achieve [...] a totally impersonal standpoint defining *the* 'moral point of view'. Care morality is about the particular agent's caring for and about the particular moral patient. Morality is not (only) about how the impersonal 'one' is meant to act toward the impersonal 'other' [...]. Not only is the self radically particularized, but so is the other, the person toward whom one is acting and with whom one stands in some relationship. The moral agent must understand the other person as the specific individual that he or she is, not merely as someone instantiating general moral categories such as friend or person in need" (Blum 1988: 476-477). There is an irreducible particularity: a particularity of the moral agent, the other and situation; 4) a moral action appropriate to a given individual is not necessarily universal or generalizable to others.

Responsiveness, the generalized, and the concrete other

We have to focus on two themes that the ethics of care underlines: the *responsiveness* of the moral subject and the difference between *the generalized other and the concrete other*. In *In a different voice* Carol Gilligan has written: "the morality of rights differs from the morality of responsibility in its emphasis on separation rather than connection, in its consideration of the individual rather than the relationship as primary" (Gilligan 1982: 20). Following the work *In a Different Voice*, in *Moral Orientation and Moral Development,* Carol Gilligan has written that ethics of care is "grounded in the assumption that self and other are interdependent, an assumption reflected in a view of action as emanating from within the self and, therefore, self-governed" (Gilligan 1995: 36). Connections among human beings are fundamental to thinking about ethics of care. What's the meaning of *taking care*? Taking care means to be able to satisfy someone's needs. The ethic of care is *responsiveness* to other's need. But who is the other? He/She can be someone in relationship with me, he/she can be someone I've never met. In both cases he/she is someone like me but different from me (see number 3 in the previous paragraph) at the same time. There's always distance between me and the other; if there wasn't that *distance*, the other couldn't be the other, he could be another myself/another me. The caring moral subject is interested in satisfying the *different-from-me-other's* needs. His/her needs can be like mine or they can be unique. The other belongs to humanity as much as I do, we are different in our common humanity, we have the same/different needs, desires, emotions, feelings, contexts, relationships. Ethics of care is focused on the uniqueness of human beings. Justice oriented ethics is an ethic of universalism: moral principles are (must be) valid for every human being in every place/time/situation. Moral action originates from an abstract rational *self* and it is addressed to abstract and generalized other. It's the Kantian model of ethics and self. According to ethics of care, the other is always a concrete other. Some years ago Seyla Benhabib distinguished between the concept of *concrete other* and the concept of *generalized other* to explain the difference between ethics of care and ethics of rights/justice. The concept of generalized other matches with ethics of rights/justice, the concept of concrete other matches with ethics of care. Seyla Benhabib has written: the "standpoint of the generalized other requires us to view each and every individual as a rational being entitled to the same rights and duties we would want to ascribe to ourselves. In assuming this standpoint, we abstract ourselves from the individuality and concrete of the other. We assume that the other, like ourselves, is a being who has concrete needs, desires and affections but that which constitutes his or her moral dignity is not that which differentiates us from each other, but rather what we, as speaking and acting rational agents, have in common [...]. The standpoint of the concrete other requires us to view each and every rational being as an individual with a concrete history, identity and affective-emotional constitution. In assuming this standpoint, we abstract ourselves from that which constitutes our commonalty. We seek to comprehend the needs of the other, his or her motivations, what he or she searches for, and what he or she desires [...]. In treating you in accordance with the norms of friendship, love and care, I confirm not only your humanity but your human individuality" (Benhabib 1987: 81). According to the vision of care the other is not "one born of others" (Benhabib 1987: 89) but a *unique individual other among unique individual others.* Ethics of justice/rights addresses *common humanity,* ethics of care addresses *humanity made of unique individuals.* We have to pay attention to the previous sentences by Seyla Benhabib and to the points previously explained. We have to pay attention to those words in relation to two fields: politics and social interpersonal sphere.

Recognition of universal human rights, recognition of identity

I will consider the communitarians' view. Twenty years ago Charles Taylor wrote: "our identity is partly shaped by recognition or its absence, often by misrecognition of others, and so a person or group of people can suffer real damage, real distortion, if the people or society around them mirror back to them a confining or demeaning or contemptible picture of themselves. Nonrecognition or misrecognition can inflict harm, can be a form of oppression, imprisoning someone in a false, distorted, and reduced mode of being" (Taylor 1994: 25). The relation between the self and the other(s) gives identity to the self and the other(s). Besides his or her self-recognition, if the self relates and compares himself or herself to the other, he or she can be himself or herself. We have to think about *recognition*. Nowadays western societies are commonly considered as needing *politics of recognition*: what does politics of recognition mean? Politics of recognition means *responsiveness* to *needs.* Firstly, the need that politics of recognition has to satisfy is the demand *of rights as universal human rights.* Without doubt the demand of rights concerns our general humanity; we could say it concerns the generalized other standpoint. All of us belong to humanity, all of us recognize ourselves participating in the same humanity that we share with others. Nowadays, besides the demand of rights, politics of recognition must to satisfy the 'demand of *identity*'. We could say this demand concerns the concrete other standpoint. This demand concerns our human individuality (as women, black people, religious minority, ethnic minority, and so on). It is based on the assumption that the self is always "encumbered"; in Gilligan's view the self is encumbered in relations, in communitarians' view (for example Michael Sandel's) the self is encumbered into ethnic group, religious group's traditions.

Conclusion

Previously we have read the words by Seyla Benhabib: "in treating you in accordance with the norms of friendship, love and care, I confirm not only your humanity but your human individuality". Caring about the other means to be able to hear/recognize other's needs, confirming his/her humanity (universal human rights), and also confirming human individuality (recognition of identity). Today making politics of recognition as recognition of identity means solving many global issues (for example: the majority of wars). Following the line of reasoning set out above we could say that making politics of recognition as recognition of identity means including ethics of care practices in social political values and actions. Politics has to care about citizens by listening to all voices voicing people's needs. Seyla Benhabib has mentioned friendship, love, and care. We need to focus on this point: friendship, love, and care are based on an emotional investment. Giving identity to others means treating them by recognizing their rights but also giving care and giving care means acting by reason and feeling, sentiments, passions, and emotions. Following the suggestion of Virginia Held, Peta Bowden, and the majority of scholars of ethics of care, we can indicate the archetypal and fundamental model of caring in *mothering*. Mothering is founded on a deep difference between two distinct individuals in a relation. Their relationship is formed of power and vulnerability. The child depends on the mother's care. The mother cares about the child. The mother is (must be) able to satisfy the child's needs. He or she is completely vulnerable while the mother has an absolute power over him or her. Silvia Vegetti Finzi has metaphorically defined this power as "something no tyrant has ever had" (Vegetti Finzi 1990: 104). It's an asymmetrical relationship. In this asymmetrical relationship the emotional investment from mother to child is the mother's motive for acting but also the reason by which she recognizes the other/the child and she partly shapes herself by recognizing herself in her child. This emotional investment is also the reason she has to reduce her power in order to form the child's autonomy, fundamental to growing up. In mothering practices, the recognition is deeply tied to and enhanced by love, feelings, and emotional investment. It is my belief that features of mothering practices could be taken as model for politics of recognition that our

western society needs to tackle many contemporary issues. We have to think about a sort of mothering deprived of a deep emotional investment. Metaphorically we have to consider the relationship between people that can give recognition and people that need recognition nearly as the relationship between the mother and her child without a deep emotional investment: an asymmetrical relation that let people involved in recognize each other. This relation gives identity both majority, owning rights and social recognition, and minorities, getting recognition. It requires majority be aware of human rights but also be aware of something else and more than human rights. It requires a *sort of care:* a care without feelings and emotions but not deprived of responsiveness. It is not possible to translate emotional investment into political decisions but all of us can certainly identify most of the others' needs: politics could start from this point.

REFERENCES

Benhabib, S. (1987), "The Generalized and the Concrete Other: Kohlberg-Gilligan Controversy and Feminist Theory", In Benhabib, S. and Cornell, D. (eds.), *Feminism as Critique. On the Politics of Gender*, University of Minnesota Press, Minneapolis, pp. 77-95.

Benhabib, S. (1992), *Situating the Self. Gender, Community and Postmodernism in Contemporary Ethics,* Polity Press, Cambridge.

Blum, L. (1988), "Gilligan and Kohlberg: Implications for Moral Theory", *Ethics*, 98, pp. 472- 491.

Galeotti, A. E. (1997), "Femminismo e teorie politiche", *Filosofia e Questioni Pubbliche*, 1, pp. 5-28.

Gilligan, C. (1982), *In a Different Voice: Psychological Theory and Women's Development,* Harvard University Press, Cambridge, MA.

Gilligan, C. (1995), "Moral Orientation and Moral Development", in V. Held (ed.), *Justice and Care: Essential Readings in Feminist Ethics*, Westview Press, Boulder.

Held, V. (1993), *Feminist Morality. Transforming Culture, Society and Politics,* University of Chicago Press, Chicago.

Kohlberg, L. (1981), *Essays on Moral Development,* Harper Row, San Francisco.

Lara, M. P. (1996), "Il concetto di soggetto morale e il femminismo"*, Parolechiave*, 10-11, pp. 287-304.

Pulcini E. (2003), *Il potere di unire. Femminile, desiderio, cura,* Bollati Boringhieri, Torino.

Pulcini, E. (2001), *L'individuo senza passioni. Individualismo moderno e perdita del legame sociale,* Bollati Boringhieri, Torino.

Taylor, C. (1989), *Sources of the Self. The Making of the Modern Identity,* Cambridge University Press, Cambridge, MA.

Taylor, C. (1994), *The Politics of Recognition,* Princeton University Press, Princeton.

Tronto, J. C. (1993), *Moral Boundaries. A Political Argument for an Ethic of Care,* Routledge, New York.

Vegetti Finzi, S. (1990), *Il bambino della notte. Divenire donna, divenire madre*, Mondadori, Milano.

GIAN PAOLO TERRAVECCHIA
University of Padua
gp.terravecchia@libero.it

SOCIAL STANCES, EMOTIONS AND THE IMPORTANCE OF FEAR[1]

abstract

The paper presents five main social stances: to refuse, to suffer, to accept, to assent and to make something one's own. They depend on kinds of relationship between an interior attitude and an exterior manifestation. The second main contribution of the paper consists in a discussion of fear and its relationships to social stances. Studying emotions helps to stress the similarities and the differences between social stances and emotions and among social stances (see e.g. rebellion and refusal). The final part of the paper tests the conclusions of the previous part by discussing Eichmann's Nazism as presented by Hannah Arendt. The paper gives an example of how ethics can be enlightened by the tools of social philosophy.

keywords

social stances, acceptance, fear, anger, Eichmann

Phenomenology and Mind, n. 11 - 2016, pp. 120-126
DOI: 10.13128/Phe_Mi-20112
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Classic moral philosophy has developed over the centuries some important and sophisticated theories to justify moral judgment, such as deontic ethics, virtue ethics, or utilitarianism. Twentieth Century ethical reflection has gone a step further, proposing a new approach: ethics can be better grounded in the light of the foundations of politics, through a discussion of the conditions of justice and social agreement. This is one specific original contribution by John Rawls, the libertarians and the communitarians. I hope that a further step, what we may call the Twenty-first Century turn, will take into account new elements offered by social philosophy. Moral philosophy, I will try to show, can better grasp its own specificity and limits by understanding that moral and social deontic norms are often interdependent and interconnected, sometimes even in conflict.

In what follows, I will present a phenomenology of social stances and I will discuss the main ways in which they relate to emotions. I will then dedicate to fear a good part of the theoretical discussion. The attempt is to show that ethics and the moral judgment can be enlightened by the tools of social philosophy.

The decision to give fear so much attention depends on the fact that it is an important emotion, for its power to give strongly compelling reasons for action. Its sometimes bold, sometimes subtle presence is very important not only for the societies evidently built on fear, but also for those just apparently free from it. It happens in fact that what it is feared does not concern just life and physical safety. The last paragraph will be a case study to test the general discussion given in the first part of the paper.

2. Social stances

An essential discussion of social stances can be developed by answering two main questions: "What are they?" and "Are they reducible to an exterior manifestation?" (see also Terravecchia 2015). Answering these questions is propaedeutic to a discussion of the relation between social stances and emotions.

A social stance is a disposition of the social agent which inclines the agent to respond in a certain way to social realities, such as requests, offers, proposals, threats, emerging social bonds. It is important to notice that the inclination here is not just a preference among alternatives, it has rather a normative dimension. Because of this, one can act in contrast to

1 I want to thank Cristina Cisotto, Paolo Del Pozzo, Roberta De Monticelli, Francesca De Vecchi, Enrico Furlan, Elena Martinolli, and Cecilea Mun for their valuable suggestions.

her stance, but in that case she has to explain her good reasons for doing that or, if need be, she has to apologise.
The social stance is a kind of "yes", or "no", or a neutral attitude, which can determine the agent's decisions, acts and further stances. For example, the treaty offered by an enemy can be accepted, or refuted. It can even be suffered if there are not acceptable alternatives. One can give her assent to the treaty, thinking that it is the best thing to do. Eventually, the agent can decide to be not just the addressee of the offer, but also to be proactive, playing a fully active role. After the decision about the social stance to be taken, everything changes. Social obligations rise for the social agent and she has to deal with them, at least until she changes the stance taken or until something else, which is essential in the situation, changes. A phenomenology of social stances shows that they are at least five: to refuse (or rebel against), to suffer (or to be subject to), to accept, to assent and to make something one's own.
Each social stance is the result of a relationship between an interior attitude and an exterior manifestation. There are three possible values for each of them: negative (to say "no"), neutral (not to say "no", which is not yet saying "yes") and positive (to say "yes"). Let us examine now each of the social stances in the light of these three values. To refuse implies a negative value from the side of the interior attitude and a negative value on the side of the exterior manifestation: the social agent does not want what is offered, proposed, etc. and she manifests it. To suffer has a negative interior attitude, but it has a positive exterior manifestation too. Here the social agent makes the best of a bad situation: though she does not like it, she grins and bears it. To accept, on the other hand, has a neutral interior attitude and a positive exterior manifestation. This is the case when the agent wants to see what happens, not having reasons to say no. The fourth case, to assent, has a positive interior attitude as well as a positive exterior manifestation. This is maybe the simplest case and is famously exemplified by the "*fiat mihi secundum Verbum tuum*" ("be it done unto me according to Thy word", Luke 1:38). Finally, the case of making something one's own does not differ in degree from the assent. It rather differs as to the distance of the agent to what has been the term of the decided stance: all the previously discussed cases entail a certain distance which is not the case here. The initiative of the other is, in the present case, fully shared in a proactive way.

3. Social stances and emotions: the threefold relationship

Social stances are dispositions. They last in time and, as said, they are the reasons for decisions and actions, and eventually to take other social stances. Laura, for example, being a friend of Kate, could have to decide whether or not to help her friend to pass an exam. Such a decision will give the background reasons to do actions such as to go to Kate's home, to study with her, to bring her a few helpful books. The disposition is a source of normativity and so social stances rule the action: as friend, Laura is obliged to assist Kate. Of course, she could decide to decline the task, but then she should show her friend that she has more urgent things to do and it would be appropriate to apologize for not helping. Explanation and apologies reveal that there are norms obliging *pro tanto*. One could point out that friendship is a social bond, not a social stance, but every social bond implies social stances and in fact, for example, one can accept or refuse a friendship or she can accept or refute her family bonds. In any case, one must take a stance: not taking any is impossible, because this would still be a case of suffering or accepting. Social normativity comes from instituted social reality or (*vel*) emerging social entities but all of them are not effective, unless they are at least suffered.
Emotions are dispositions too (this remains true, even if moods are more stable as dispositions, see Deonna, Teroni 2012: 4), but social stances as such are not emotions. The ways in which social stances are related to emotions is at least threefold. Social stances (1) can be contingently connected to emotions, (2) they can be typically accompanied by emotions and they can be (3) causally related to emotion. An example of a contingent connection is seen

in the case of a girl who has received a coloured and funny love declaration, emphatically performed by a clown on the street. The situation explains why she is amused, an emotion which of course is not dependent on her smiling refusal. Indeed, refusing someone's love declaration, as such, is not connected with amusement. However, the same girl would typically feel joy, probably mixed with embarrassment, if the clown were her boyfriend crazily though seriously asking for her hand. In this case, her assent would be typically accompanied by joy. Other examples of typical though not necessary connections between social stances and emotions are indignation, associated with refusal when what is refused has to do with important disvalues, and anger, associated with rebellion. Here the emotions help to shed light on the differences between rebellion and refusal. Such differences are not evident from a strictly cognitive perspective, both being a "saying no". On the other side, apathy is typically connected with acceptance. Finally, to exemplify the third case, emotions are the reasons for assuming social stances: one can give her assent to join a political party because she likes its leader, or a girl can give her assent to marry the clown, because she loves him. It is clear that the causal relation is not excluding freedom, it rather shows the reasons for its exercise.

4. Fear and social stances

Fear plays a central role in the history of political philosophy. Thomas Hobbes, as known, in his *Leviathan* (1651) indeed considers fear as the root of political life: it is fear that pushes people to accept the social contract (see ch. 17 and ch. 20 where he distinguishes between the reciprocal fear and the fear of the Sovereign). It is disputable and indeed it is widely discussed in political philosophy whether Hobbes is right on this within the political theory (see e.g. Blits 1989, Debrix 2009, Ginzburg 2008, Goodin, Jackson 2007). In social philosophy however, or at least in the phenomenology of social stances, the role of fear is clear. Fear can be a reason for the first two social stances (to refuse and to suffer). About the third stance, acceptance, fear should be at least accompanied by other positive reasons or emotions to reach a balanced, neutral, interior disposition. Finally, fear cannot be presented as a positive motivating reason in the last two social stances, to assent and to make something one's own. It may seem that someone could make her own something because of fear, but if fear is the reason of her stance, then she is just acting *as if* she were making it her own (having the possibility, it may be interesting here to discuss the special case concerning self deception). These two last stances require a positive interior attitude and this excludes fear as a direct reason. All this also shows that the social stances on which proactive behaviour and social communities are largely grounded do not have fear as direct reason. Therefore, fear is not sufficient to give an account of the social reality, as shown by a phenomenology of social stances.

The relationship between social stances and fear is threefold: fear can typically accompany one social stance as we will see; it is causally related to another social stance and, finally, it is contingently connected with a third one. As for the first case, to suffer is typically connected with fear, though of course not exclusively. From ancient times to Twentieth Century totalitarianisms fear has been an *instrumentum regni*. In ancient societies, assent was considered unnecessary and refusal was made impossible, or at least very dangerous, acceptance was difficult to be gained by the power, so to suffer was the typical option for the people. Rebellion was terribly punished with a twofold result to destroy the stronger opponents and to give an example: entire cities robbed and razed, people enslaved after many atrocities or brutally killed. All this was meant to create the background common knowledge that to refuse the requests of the power is extremely dangerous. It is now easier to understand Machiavelli's reasonable advice to the prince (*The Prince*, 1532): "it is much safer to be feared than loved" (ch. 17). Interestingly enough, Machiavelli's further advice is to avoid that fear becomes hatred. We can understand why: hatred is strong enough to motivate a rebellion, as fast as it is doable. Hatred, indeed, gives very strong reasons to revolt notwithstanding fear.

After many attempts, blood spread and tears, the human kind has found, through institutions (Montesquieu's doctrine of separation of powers is a classical example), forms of collective life not ruled by fear for life and for essential values (the so called human rights).
Fear can be causally related to refusal: one can refuse the request to take a flight, because she has an irrational fear of flying, even if she knows that an airplane is the safest means of transportation. Let us also consider a man who refuses a puppy as gift because he fears dogs: as a child he was badly bitten. From a rational point of view, he can explain his fear, but still the reason for his refusal is an instinctive emotion that sounds odd, watching the cute puppy. Yet, fear sill motivates quite reasonably the stance of not having anything to do with dogs. Here the emotion becomes the reason for action. Political history gives many examples of causal relation of fear with social stances. French revolution history maybe presents the most famous example: The Great Fear (17 July - 3 August 1789), a popular revolt in the countryside.

5. Fear: its contents and its relevance in action

There are still two important topics to be discussed about fear within the field of social philosophy: its contents and its relevance in action. As to contents, it is necessary to distinguish at least between four main general contents: the fear for life and for one's own physical safety; the fear for one's honour; the fear for the agent's properties, and finally, the fear for the safety of other people, especially of those with whom the social agent is bound. Of course, the first content of fear is at least *prima facie* much more compelling than the others, but one should not underestimate these, especially the second and the fourth. Social life could be so much negatively conditioned by the loss of honour that it may well happen that the agent, after loosing it, finds no reason to live so that the final result of the first two cases is the same. About the fear for those with whom the social agent is bound, it can be an even stronger reason for acting than fear for one's own safety: the agent who may accept the sacrifice, may not be ready to sacrifice her loved ones.
Even in societies that would be free from physical violence, the fear system would not be necessarily absent. There are forms of pressure that have to do with threatening the honour. These forms of violence can be hard to be detected, since they tend to be subtle, working on the background, from the "not said". Nonetheless, they are effective and they can deeply influence, and eventually manipulate, people.
A discussion of the contents of fear helps to understand how much it is relevant and from this the levels of responsibility of the social agent. Understanding the connection between fear and actions is essential to morally evaluate such actions, because the presence of fear may help to scale down the responsibility of the agent. This is true for a single act, as well as for social dispositions. For example, a man obliged to do something bad by a gunman threatening his life is judged with indulgence by the law and from a moral point of view. Here the emotion plays an important role, showing the relevance of the interior attitude in taking a stance. Moreover, the presence of the emotion of fear, if detected, helps to understand the origin of the action for its deep compelling roots and the limits of the free act.

6. Testing the theory

It is useful to test a new conceptualisation, to show if and how far it explains the phenomena, i.e. if it helps to better understand the facts and the values involved in real life situations. For this reason, I will now discuss an exemplar case to see if the theory works and if it is useful for a better understanding of the social reality. I take as an example the notorious case of Adolf Eichmann, as presented by Hannah Arendt in her famous *Eichmann in Jerusalem* (Arendt 1963). For the purpose of the present discussion, it is enough to test if the proposed theoretical framework works well for Arendt's Eichmann, so to say.
Nazism gave the German people positive reasons to be embraced, such as its fight against unemployment and against hyperinflation, its defence of national values, the perspective of a

glorious future for the country. It provided also negative reasons, e.g. the fear of communism or the idea that "without us disorder and chaos will reign". Considering the hard times for Germany in the Thirties, from an internal and an international point of view, and considering the weakness of its institutions, the positive reasons could have been easily turned into negative. In this case they would sound as the fear of unemployment, the fear of being a humiliated country that has lost its values. All this gave many Germans the reasons to suffer the Nazis presence or to accept their methods. Others were ready to refuse Nazism, but not having the power to change the situation by fighting as individuals, they had to suffer the events. The fear for one's safety and sometimes for one's life prevented almost any resistance and led the vast majority of Germans to suffer or to accept Nazism.

In this historical context, Eichmann decided to join the S.S. accepting a proposal. Here it is Arendt's page about the episode: "Kalterbrunner had said to him: Why not join the S.S.? And he had replied, Why not? That was how it had happened, and that was about all there was to it" (Arendt 1963: 31). Eichmann initially did not give his assent to Nazism, as Arendt well explains: "He had no time and less desire to be properly informed, he did not even know the Party program, he never read *Mein Kampf*" (*ibidem*). His frustrating life of a mediocre man of the lower middle class, a condition without prospects, pushed him to say a kind of "let us see what happens". If this is his starting point, after a while the propaganda, his sincere admiration for Hitler and his desire to pursue a career, led him to give his full assent to Nazism and to follow orders scrupulously. Here one can see how the party worked, forging his interior disposition: "He was perfectly sure that he was not what he called an *innerer Schweinehund*, a dirty bastard in the depths of his heart; and as for his conscience, he remembered perfectly well that he would have had a bad conscience only if he had not done what he had been ordered to do – to ship millions of men, women, and children to their death with great zeal and the most meticulous care" (Arendt 1963: 23).

He was not acting because of hate against Jewish. He somehow considered himself as an idealist. Here is the explanation given in the book:

"An «idealist» was a man who lived for his idea – hence he could not be a businessman – and who was prepared to sacrifice for his idea everything and, especially, everybody. When he said in the police examination that he would have sent his own father to his death if that had been required, he did not mean merely to stress the extent to which he was under orders, and ready to obey them; he also meant to show what an "idealist" he had always been" (Arendt 1963: 40).

Hate would have explained a stance of refusal of the Jewish people. Eichmann in many occasions and with many examples tried to show that this was not his position up to the point that when he was told about the Final Solution, these were his feelings: "I now lost everything, all joy in my work, all initiative, all interest; I was, so to speak, blown out" (Arendt 1963: 29). Here we see a case in which emotions accompany but not determine the social stance. We also see a conflict of the social norm (obey orders) with the moral one (do not harm the innocent) and we know how he solved the dilemma tragically.

Eichmann's position, during the process, would have been better if he had demonstrated that he was under the influence of fear for his life, or for the life of his family, when complying with Nazi's orders. Arendt, for example, writes about Himmler's "justified physical fear of Hitler". His fear, however, did not prevent Himmler to stop the Final Solution. He hoped in this way to get a merciful treatment after the War. Eichmann could have simply followed Himmler and the "moderate wing" of the S.S., but he preferred to be faithful to the *Führer*, challenging Himmler's orders. After all, neither his free joining the S.S., nor his doing his job "with great zeal" had anything to do with fear. For these reasons we can say that he did not suffer, nor he merely accepted Nazism. He gave his full assent to it and this was considered unforgivable by the jury. The absence of fear, where expected as a discharge, is an aggravating factor.

7. Conclusions

The first part of the paper presents the five social stances: to refuse (or to rebel against), to suffer (or to be subject to), to accept, to assent and to make something one's own. They are a result of a relationship between an interior attitude and an exterior manifestation in a way that characterizes each of the stances. The social stances are not emotions and the first main contribution of the paper was to discuss the threefold relationship between social stances and emotion: contingently connected, typically connected and, finally, causally related. The second main contribution was given through a discussion of fear, by showing its relationships to social stances. Fear was discussed taking into account its four main general contents (the agent's life and physical safety, the agent's properties, the honour, and finally the safety of others, especially the loved ones) and its relevance in action.

Among the secondary and still relevant results of the paper, it was possible to notice that discussing emotions helped to stress some aspects otherwise not evident from the mere discussion of the social stances: rebellion and refusal, have in common their being a "saying no", but they differ because rebellion, but not refusal, is typically, though not necessarily, connected to anger. It was also important to stress a common feature of social stances and emotions: they all are dispositions.

Through its specific contributions the paper tries to give a wider contribution by showing that ethics and the moral judgment can be enlightened by the tools of social philosophy. In fact, a case such as Eichmann's trial was discussed in depth in the last part of the paper. Social philosophy reveals a normativity which interacts and sometimes enters in conflict with the normativity of ethics. In this case we have the social-moral dilemmas such as: "Am I obliged to obey the orders, or must I follow the ethical rule not to harm the innocent?". Emotions add a further level of complexity to this field: their compelling power may interfere with the two normative domains adding some more issues. What they do, or miss to do, is relevant to judge the action.

REFERENCES

Arendt, H. (1963), *Eichmann in Jerusalem. A Report in the Banality of Evil*, Penguin Books, New York, 2006;

Blits, J. H. (1989), "Hobbesian Fear", in *Political Theory*, 17, pp. 417-431;

Debrix, F. (2009), "Nothing to Fear but Fear: Governmentality and the Biopolitical Production of Terror", *International Political Sociology*, 3, pp. 398-413;

Deonna, J. A., and Teroni, F. (2012), *The Emotions: A Philosophical Introduction*, Routledge, London;

Ginzburg, C. (2008), *Fear Reverence Terror. Reading Hobbes Today*, European University Institute, Firenze;

Goodin, R. E., and Jackson, F. (2007), "Freedom from Fear", *Philosophy and Public Affairs*, 35 (2007), pp. 249-265;

Machiavelli, N. (1532), *The Prince*, English translation: http://www.gutenberg.org/ebooks/1232;

Terravecchia, G. P. (2015), "A Phenomenology of social stances", *Phenomenology and Mind*, 9, pp. 108-115.

SESSION

SESSION 3

EMOTIONS, LANGUAGE, AND HATE SPEECH

ROBIN JESHION
University of Southern California
jeshion@dornsife.usc.edu

SLUR CREATION, BIGOTRY FORMATION: THE POWER OF EXPRESSIVISM[1]

abstract

Theories of slurs aim to explain how – via semantics, pragmatics, or other mechanisms – speakers who use slurs convey that targets are inferior persons. I present two novel problems. The Slur Creation Problem: How do terms come to be slurs? An expression 'e' is introduced into the language. What are the mechanisms by which 'e' comes to possess properties distinctive of slurs? The Bigotry Formation Problem: Speakers' uses of slurs are a prime mechanism of bigotry formation, not solely bigotry perpetuation. With a use of a slur, how are speakers able to introduce new bigoted attitudes and actions toward targets? I argue that expressivism offers powerful resources to solve the problems.

keywords

slurs, epithets, pejoratives, semantics, racism

Phenomenology and Mind, n. 11 - 2016, pp. 130-139
DOI: 10.13128/Phe_Mi-20113
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Slur Creation, Bigotry Formation

Slurring terms are pejorative expressions that target individuals on the basis of race, ethnicity, religion, sexual orientation, nationality, socioeconomic status, occupation, and various other socially important properties. They are tools of subordination and their use a threat to human dignity. What do such expressions mean? And why are uses of them so extraordinarily offensive and so destructive to the groups and individuals they target?

My overarching perspective on slurs is guided by the idea that slurs function to dehumanize. When someone uses a slur, he *somehow* conveys that the targets have lesser standing as human beings. They are inferior, undeserving of the full respect we accord to persons *qua* persons. To dehumanize, slurs need not convey that targets are subhuman, creatures wholly undeserving of all respect. They need 'only' convey that targets are beneath the rest, possessing lower status along the moral domain, broadly construed.

In this paper, I will assume, not argue for, this guiding idea. Most theorists make the same assumption, at least implicitly[2]. Indeed, one of the fundamental common problems and thus goals of philosophical theorizing on slurs is to explain *how* – via semantics, pragmatics, or some other mechanisms – speakers who use slurring terms convey that targets are inferior persons *qua* persons. Let's call this the *Dehumanization Problem*.

One of my primary aims is to present two new problems confronting theories of slurring terms. Both are importantly connected to, yet distinct from, the *Dehumanization Problem*. Another is to argue that expressivist theories of slurring terms possess excellent resources to solve these two new problems. While I do not maintain that other theories cannot solve the two problems, I illustrate why several are limited in their ability to do so.

The first problem is the *Slur Creation Problem:* How do terms come to be slurring terms? An expression is introduced into the language. How exactly does it come to possess the properties distinctive of slurs? At time t, expression 'e' is introduced into the language. At some later time t', 'e' is a slur, possessing the semantic, pragmatic, or social properties (the theory) deemed

1 I presented material from this paper at the San Raffaele Spring School *Emotions, Normativity, Social Life Workshop* in June 2016 and at the Dubrovnik Workshop on Philosophy of Language in September 2016. My thanks to both audiences for their interesting, challenging questions.

2 The agreement is typically not stated in the terms I employ, as slurs functioning to *dehumanize*, but rather as functioning to derogate others or convey targets as unworthy of full respect. Cf., Hom (2008) on derogatory force: epithets forcefully convey hatred and contempt of their targets. Even if there are substantive differences reflected in the formulation, they will be irrelevant for the purposes of this paper.

definitive of slurs. How did that happen? What are the mechanisms by which it attained the properties of slurs?
This problem is in many respects akin to another concerning meaning change, the problem of appropriation. At some time t, expression 'e' is a slurring term, possessing the semantic, pragmatic, or social properties (the theory) deemed definitive of slurs. At some later time t', 'e' either lacks the properties definitive of slurs or has come to be ambiguous, with one of its new meanings lacking the properties. Just as theories of slurs ought to be capable of shedding light on how slurring terms' derogation and, more broadly, their offensiveness can come to be *neutralized* through social processes of appropriation, so too should they explain the means by which a term comes to secure its particular slur-characterizing properties in the first place.
The second problem is the *Bigotry Formation Problem*: Speakers' uses of slurring terms are a prime mechanism of bigotry formation, not solely bigotry perpetuation. In using a slurring term 'e', a speaker is capable of contributing to the formation, including the initiation, of the hierarchical group structures (in-groups and out-groups) that partly constitute social institutions of bigotry. Following Hom, let's characterize institutions of bigotry as composed of bigoted ideologies and bigoted practices toward members of the group. Bigoted ideologies are sets of bigoted beliefs and others attitudes toward and about a particular group. Bigoted practices are systematic ways of acting toward members of the group that are rooted on and rationalized by reference to the bigoted ideology. These include but are hardly limited to intentional systematic exclusion from positions of power, being accorded unequal treatment or less protection by law, all varieties of violence toward group members for being group members, from schoolyard bullying and humiliation to genocide. The problem of bigotry formation, then, is that with a use of a slur, a speaker is capable of introducing new bigoted beliefs, attitudes, and actions toward members of the target group and our account of slurring terms must be capable of explaining this fact[3].

2. Expressivism

My favored analysis of the semantics of slurring terms is a specific version of hybrid expressivism[4]. Here I provide a bare-bones sketch of the view. It pertains specifically to literal uses of slurring terms. Slurring terms function semantically in roughly the same way that their neutral counterparts function when accompanied by contemptuous intonation ([1b], [2b]) or fronted by certain expletives or contempt-expressing adjectives ([1c], [2c]).

[1a] Jake is a Kike.
[1b] Jake is a *Jew*^{C}.
[1c] Jake is a dirty Jew.

[2a] Is that bar full of faggots?
[2b] Is that bar full of *homosexuals*^{C}?
[2c] Is that bar full of goddamn homosexuals?

There are three separable components to slurs' semantics, what I call the *group-designating*, *expressivist*, and *identifying* components.
The group-designating component: slurring terms designate a particular group, the very

3 My reflection on the problems of slur-creation and bigotry-formation was inspired in many ways by Tirrell's excellent (2012).
4 Details are in Jeshion (2013, 2016a). Others who adopt versions of hybrid expressivist semantics include Copp (2001), Potts (2005), Saka (2007), Richard (2008).

group that their neutral counterpart designates, if in fact the slur possesses a neutral counterpart. (Some don't, as we'll see momentarily.) It is the only component that contributes to determining the truth-conditional content of truth-conditionally evaluable utterances containing slurs. Just as [1b] is true just in case Jake is Jewish, so too is [1a] true in just such a circumstance[5]. Because it encodes exactly what its neutral counterpart does, the group-designating component contributes nothing to solving the Dehumanization Problem.
The expressivist component: slurring terms are vehicles for expressing – not stating or declaring – speaker's attitudes. With a use of a slur, a speaker expresses his attitude of contempt for members of a socially relevant group G on account of their belonging to G or having a group-defining property *g*. Along this dimension, slurs are analogous to other expressions that are, as a matter of convention, used to express speakers' affective attitudes. "Hooray!" is used to express a speaker's happiness, often at triumph; "holy crap!" is used to express a speaker's astonishment or fear. Neither one encodes semantic descriptive contents like "I am happy!" or "I am astonished!". Likewise, utterances of [1a], [2a] do not encode the semantic descriptive content encoded in [1d], [2d], respectively.

[1d] Jake is Jewish and is therefore worthy of contempt
[2d] Is that bar full of homosexuals who are contemptible on account of being homosexual?

Instead, the speaker simply *expresses* the affectively laden evaluative attitude toward those designated by the slur's neutral counterpart, just as he does when sincerely expressing his attitudes by inflecting "Jewish", "homosexual" with a tone of contempt, as in [1b], [2b]. Understanding the expressivist component requires understanding the moral-psychological structure of contempt. Contempt is an affective attitude that can be manifest by a whole range of emotions (aversion, hatred, disgust, condescension, pity) and dispositions (to scorn, mock, humiliate, turn away, withdraw, and disregard). As such, it is an emotion, yet one that need not be accompanied by any particular behavioral manifestations or discrete feelings. It is also an evaluative attitude. Contemptuous regard involves ranking another person as low in worth along the moral domain on a certain basis. Consequently, in uttering [1a], by the expressivist component of the semantics, the speaker expresses his affective attitude that Jake and Jews more broadly rank low in worth *qua* persons on the basis of being Jewish.
The identifying component: in using a slur, a speaker classifies and represents the target in a way that aims to be identifying, that aims to specify *what the target is*. In [1a], just as in [1b] and [1c], the speaker conveys that being Jewish is a fundamental negative characteristic-defining feature of the targets, here Jake and Jews more broadly. This identifying reductive classification is an aspect that follows from the expressivist component because of the role of the basis within the moral-psychological structure of contemptuous regard. Because contempt (unlike hate, say) has a basis, contemptuous regard for another involves taking the properties of that basis as fundamental to the target's identity as a person[6].

5 This claim is obviously highly controversial. I attempt to defend it in Jeshion (2016b).
6 Although the identifying component follows from the expressivist component (as a matter of the psychology of contempt), I dignify it as a separate component of the semantic of slurs because it is separately *deniable*. In response to [1a], Jake could, for instance, agree that being Jewish is fundamental to his identity while denying the contemptuous regard. I defend the first point about the psychological structure of contempt in Jeshion (2016a).

3. The Power of Expressivism

Expressivism provides a straightforward answer to the Dehumanization Problem. Speakers who use slurring terms convey that targets are inferior persons *qua* persons by virtue of the semantics: in sincerely using a slur with full understanding, speakers *express* their contempt for the targets, an affective attitude that the target ranks low in worth as a person. Furthermore, because the expression of contempt is semantically encoded, slurs possess a specific communicative function: to dehumanize in the sense of performing an action of *treating* targets as persons unworthy of full respect. Second-person uses, *callings* ('You kike!', 'Faggot!'), constitute direct acts whose function it is as expressions of contempt to induce shame in targets, compel a self-representation or self-assessment as someone lesser, unworthy, undeserving of respect. Third-person uses constitute acts whose function it is to lower the target's worth in the eyes of others, and consequently to define their social standing *as* lesser[7].

Expressivism can also offer a plausible solution to the Slur Creation Problem. In my view, there may be multiple distinctive mechanisms that contribute to slur creation. Understanding how terms come to be slurring terms demands a full investigation of them, one resting on both conceptual analysis and quasi-empirical research into the dynamics of meaning-formation of various particular slurs. Here I spotlight what I regard as the most basic and plausible mechanism by which terms come to be slurs. I call it *Contempt Crystallization*.

Suppose that at time t, an expression 'e' is newly introduced into the language and is used to designate members of a certain group G. To mock, humiliate, or otherwise put down or signal disrespect toward members of G, speakers who use 'e' accompany its utterance with canonical behavioral manifestations of contempt. Cognitive psychologists have identified two primary behaviors cross-culturally expressed and discernable as contempt. One is contemptuous intonation. The other is the so-called 'unilateral lip curl', the facial expression in which one corner of the lips is retracted and raised, and eyes are slightly narrowed. Both are easily detectable and identifiable as expressions of contempt[8]. Because the manifestation of such intense affect often engenders transmission of feeling and mimicking reactions in non-targeted interlocutors disposed to similar social stances, others will come to use 'e' in the same fashion, manifesting contempt themselves. Over time, the contemptuous regard that is behaviorally manifest comes to 'rub off' on 'e' so that it eventually becomes crystalized in 'e': at some t', 'e' conventionally encodes the speaker's contemptuous regard even in the absence of obvious behavioral manifestations. If the community continues to use 'e' for and only for the same group G that the initial users aimed to designate with 'e', the term comes to be a slur for G. According to expressivism, this systematic use by speakers to designate and indicate

7 Dehumanization often occurs, in addition, through specific features of the context of utterance. For instance, with a use of "faggot" in a particular context, a derogating group-stereotype may be activated, thereby contributing a subsidiary *pragmatically-induced* means by which gays are dehumanized. I discuss many distinct pragmatic sources of offensiveness in Jeshion (2013). Quite generally, on my view, to explain all the phenomena slurs exhibit, the expressivist semantics requires supplementation from a rich pragmatics. The semantic theory is necessary to account for what is constant across literal weapon uses of different slurs, and to explain how they differ from neutral counterparts.

8 Researchers have established that like many other basic emotions, contempt is manifest in specific *acoustic patterns*, its affective prosodic profile. These are construed as a function of the level, range, and contour of its fundamental frequency, its amplitude, intensity range, speech rate and various other temporal phenomena. To test the capacity to discern the acoustic pattern of contempt as contempt, researcher employ trained actors to utter neutral or nonsense sentences in the acoustic pattern of the emotion. Listeners are able to distinguish it as contempt with accuracy well beyond the statistically significant, comparable to that of many other negative emotions. Cf., Banse and Scherer (1996). For research supporting the unilateral lip curl as a universal expression of contempt, cf., Ekman and Heider (1988) and Matsumoto and Ekman (2004).

contempt toward the group is sufficient for 'e' to become a slur[9].

As a basic template for slur-introduction, *contempt crystallization* contains several notable features. First, unsurprisingly, it is akin to the introduction into the language of other expressives. According to Merriam-Webster's Dictionary, the first known use of "woohoo!" was in 1981. Imagine its initial uses. We can plausibly suppose that the term was first accompanied by the speaker's prosodic or facial expressions of excitement or delight. The speaker of the utterance employed a compelling sound, one useful for efficiently connoting excited pleasure. People liked it. Imitation proliferates. Communal correlation of term with speaker's affect ensues. Eventually, the affect becomes crystalized with the term, allowing "woohoo!" to encode speaker's excitement *sans* demonstration of affect. Conventionalization. This illustrates the obvious: new expressive meanings are grounded in the expression of attitudes. Contempt crystallization is just a special case of *affect crystallization*.

Second, contempt crystallization does *not* depend upon the existence of a simple neutral counterpart expression with which the slurring word *contrasts* with respect to its register or with respect to being marked. The feature is attractive because many slurring terms in fact fail to possess neutral counterparts. Consider "hillbilly", "redneck", "white trash", "dago", "wog", and "gook". The first two are slurs used primarily by northern Americans and wealthy southerners to refer to uneducated, poor rural people living primarily in the Ozarks and Appalachia. "Dago" is used to refer wholesale to Italians, Portuguese, and Spaniards, "wog" by British and Australians to refer to anyone of Middle Eastern, south Asian, eastern European, and Mediterranean descent. "Gook", though primarily used today to refer exclusively to Koreans and Vietnamese, since the turn of the twentieth century, the expressions was the slur of choice in the US military applied indiscriminately to natives in US occupied counties (Nicaraguans, Costa Ricans, Filipinos, Korean, Cambodians, Vietnamese) and all enemy soldiers.

Third, contempt crystallization does not depend upon the existence of disparate opinions in the community about the worth of the slur's targets. Everyone could accept their standing as lesser, even the targets themselves could assume a false-consciousness. Our analysis of slur introduction thereby allows for the existence of slurs within insulated like-minded communities[10].

Fourth, contempt crystallization not only explains how terms come to be slurs but also the *rapidity* and *ease* with which they can do so. To use the contemporary jargon of cognitive psychologists, terms can rapidly morph to slurs by means of emotional contagion.

Fifth, in contempt crystallization, the mechanism by means of which the term secures its referent need not line up exactly with the mechanism through which it comes to encode contempt. This affords our analysis considerable flexibility to explain various mechanisms of meaning acquisition. Consider 'kike'. Its earliest documented use is 1900. Two accounts of its origins predominate. On one, due to Leo Rosten, the word was coined on Ellis Island. Illiterate Jewish immigrants needing to sign entrance documents would use a circle 'O', refusing to sign with the customary 'X' that they associated with the Christian cross. As the Yiddish word meaning "circle" is "kikel", immigration inspectors used the term to shepherd people through the documentation process, and eventually began calling anyone who signed with an 'O' a kike. Here, I hypothesize that the expression "kike" was initially accompanied with manifestations of contempt by the non-Jewish inspectors, employed to refer to Jews more broadly and its

9 In maintaining that this account of contempt-crystallization is sufficient for slur-creation, I am not denying that numerous other social factors are typically present and can contribute to slur-creation.

10 Hom and May (2016) make the important point that slurs can exist in such communities.

intended referent stuck. On an alternative account, assimilated German Jews introduced the term as a derogative expression exclusively for eastern European and Russian Jews. Since their names often ended with the sound 'ki' ("Piatigorsky", "Brodsky"), "kike" accompanied by contemptuous intonation initially served as a put down for a subgroup of Jews. Over time, the term was picked up by non-Jews, applied wholesale to the group as an expression of anti-Semitism. Both account are plausible. (I take no stand on the actual historical development of 'kike'. Both are plausible *patterns* for slur-formation.) And both are explicable on an expressivist semantics that sharply separates the group-designating component from the expressivist and identifying components.

Finally, our analysis is *organic* because the primary mechanism of slur creation – the behavioral exhibition of contempt – is precisely what explains why slurs dehumanize. That is, the main resource in our solution to the problem of slur creation itself provides a natural analysis of slurs as tools of dehumanization.

What of the bigotry formation problem? What resources does expressivism possess to demonstrate how slurs contribute not just to the perpetuation of bigotry but its inception as well? The fact that expressivism allows for an organic solution to the problem of slur creation itself demonstrates how slurs contribute to bigotry formation. At all stages in the process of contempt crystallization, speakers exhibit their contempt for targets, a key mechanism for the rapid transfer of contemptuous attitudes toward the group. A wide range of psychological research establishes that manifesting *attitudes* toward others produces visceral mirroring reactions and responses in interlocutors. It can thereby contribute to explaining how, in ripe opportune social circumstances, contempt *goes viral*[11].

Furthermore, the nature of contemptuous regard discloses how its expression in slurs contributes to the formation of hierarchies partly constitutive of institutions of bigotry. Because uses of slurs encode dehumanizing attitudes and execute acts that aims to shame and lower the social status of targets, users of slurs contribute to the formation of bigoted attitudes and the hierarchical social structure (in-groups out-groups) constituting our most basic moral categories, those deserving full respect, those that do not.

4. Slur Creation, Bigotry Formation: Alternative Theories of Slurs

Can other theories solve the dual problems of slur creation and bigotry formation? I will not argue that they confront an insurmountable obstacle. I wish only to show that certain non-expressivist theories of slurs encounter a common difficulty. To put the point crudely, all of them analyze what slurs are and how they dehumanize by piggybacking off of pre-existing bigoted social institutions. Yet in appealing to bigoted social institutions to characterize slurs, all have significantly limited their resources for explaining how those very terms came to be slurs and how they contributed to constructing the bigoted institutions they are characterized in terms of.

Furthermore, all of them, and indeed other theories of slurs as well, ought to confront contempt crystallization. Its possibility and, I maintain, plausibility as an effective mechanism for slur creation puts pressure on non-expressivist theories. If contempt crystallization is recognized as a plausible mechanism for slur creation, why not then adopt a full-fledged expressivist semantics of slurs? After all, the main resources in the account of slur creation are the same as in an expressivist semantics. If contempt crystallization is deemed implausible, one ought to say why, especially in light of its construction from research on the emotions and standing as just a special case of affect crystallization.

The views I consider are Hom's (2008) descriptivist analysis of slurs as semantically

11 Cf., Hatfield, E., Caciopp, J., Rapson R. (1993), (1994) for early work on emotional contagion.

encoding group stereotypes enshrined in bigoted institutions; Anderson and Lepore's (2013) minimalist semantics coupled with an analysis of slurs' offensiveness construed entirely in terms of prohibitions on slurs' use; Nunberg's (2016) minimalist semantics coupled with his pragmatic affiliationism that construes slurs as marked terms signaling bigoted ideologies and institutions via generalized conversational implicatures[12]. Although there are significant strengths and problems with each view, I focus here almost exclusively on the common concern regarding their ability to explain slur creation and bigotry formation.

On Hom's view, slurs piggyback on existing social institutions via their semantics. They encode a semantic descriptive content referencing properties about the group contained in the bigoted social ideology together with a discriminatory set of practices regarding how the group ought to be treated, where the ideology purports to justify the practices. For instance, "Chink" according to Hom "expresses a complex socially constructed property like: ought to be subject to higher college admissions standards and ought to be subject to exclusion from advancement to managerial positions, and...because of being slanty-eyed, and devious, and good at laundering, and...all because of being Chinese." Hom (2008) 431. The trouble regarding bigotry formation is straightforward: how could utterances of "Chink" contribute to forming, and especially *initiating*, racist ideologies and practices toward Chinese persons if its meaning encodes them? True, utterances could *perpetuate* the already enshrined racism encoded in the word. But it is doubtful that *via its semantics* "Chink" could shape the racists ideology[13].

The problem of slur creation is that there appears to be no natural process by which a term is introduced into the language and comes to have the semantic descriptive content Hom offers for "Chink". Because Hom treats slur-content as individuated externalistically, he might align slur creation with natural kind term introduction. "Water" has its reference fixed externally, to designate *that* kind of stuff. So too for slurs. But by itself, this lacks plausibility because both "Chink" and "Chinese" will be introduced in the same fashion, to refer to *those* folks, and there is no reason why the slur and only the slur will come to possess the pejorative stereotype-containing externalist content. To make a case for how "Chink" evolved to become a slur, we need to understand why it, but not its neutral counterpart, ends up incorporating a content that implicitly references racist ideologies.

On minimalist semantics to which Anderson and Lepore and Nunberg subscribe, slurs are synonymous with their neutral counterparts. Anderson and Lepore advocate *prohibitionism*, a version of minimalism according to which slurs possess no special *linguistic properties* at all. Slurs are both semantically and pragmatically on a par with their neutral counterparts. What distinguishes the two is not to be explained by semantic descriptive content, nor by conventional implicature, nor by a particularized or generalized conversational implicatures. Everything about slurs' distinctive patterns of offensiveness is instead explained by the *extra-linguistic* fact

12 For lack of space, I bypass considering Bolinger's and Camp's accounts. If Bolinger is read as adopting semantic minimalism (she leaves this somewhat unclear), then what I say about Nunberg's view applies to hers. I suspect that because Camp analyzes slurs in terms of pre-existing perspectives, her accounts confronts the same problems as the others. Cf., Camp (2013), Bolinger (2015).

The views of Hom, Anderson and Lepore, and Nunberg are all striking insofar as they aim to explain all the phenomena slurs exhibit by appealing to a *single feature*, a descriptivist semantics, prohibitions on slurs' use, and affiliations with social ideologies of the slurs' provenance, respectively. My view by contrast recognizes multiple distinct varieties of sources to account for the phenomena. Some is explained by the semantics, some by pragmatics, some simply by reference to what speakers know about their social environments. Saka's discussions of slurs also draw upon multiple resources. Cf., Jeshion (2013b, 2016), Saka (2007).

13 Of course, one could supplement the ideology by saying things like "Chinks are so weird" but one could do the same with the neutral counterpart. The central intuition is that whatever is distinctive of slurs *itself* contributes to the bigotry formation. For Hom, it must be through slur's semantics.

that "their uses are prohibited, and so, they offend those for whom these prohibitions matter" (Anderson and Lepore 2013: 18). Here, the status of being a prohibited word is purely social: "slurs are prohibited words not on account of any content they communicate, but rather because of relevant edicts surrounding their prohibition" (Anderson and Lepore 2013: 18).

One of the most serious concerns for this view is its inability to account for how slurs derogate. The harm engendered by breaking taboos is radically different in kind from the harm engendered by spewing slurs. This concern is deepened when prohibitionism confronts the bigotry formation problem. Social prohibitions on their use, said to pinpoint what is offensive in slurs, simply cannot themselves explain anything about how slurs contribute to bigotry formation. Breaking prohibitions may cause offense by disrespecting those who enact and promote the prohibitions, but such disrespect is radically different from the societal bigotry formed by uses of slurs. Slur creation poses an equal challenge. For Anderson and Lepore, the initiation of a prohibition is what changes an ordinary term that references a group into a slur. Yet this analysis of slur origins is suspect: "Nigger" and "Chink" and "kike" were slurring terms far longer than blacks and Chinese and Jews had sufficient power in society to "banish" their use. Slurs come to be prohibited precisely *because* they *had been* functioning to dehumanize.

Nunberg's version of minimalism, what he calls *affiliationsim*, explains slurs' power to derogate and offend by systematic pragmatic processes that result in a 'routinized' conversational implicature. These ventriloquistic implicatures, as Nunberg calls them, can be executed when a speaker confronts a choice in referential expression. She possesses two synonymous terms, yet one, the slur, is marked and associated with the bigoted attitudes of those who regularly use it and its provenance; the other, the slur's neutral counterpart, is unmarked, the preferred term in 'polite' society. A speaker who uses a slur flouts the manner maxim by pointedly choosing the marked expression, which results in the speaker affiliating herself with the attitudes of the bigoted group that uses it[14].

Does affiliationism possess the resources to explain how slurs contribute to bigotry formation, including its initiation? The primary worry is that while speakers are able to generate the effects of slurs by affiliating themselves with pre-existing bigoted ideologies and groups that employ the term, their capacity to do so is executed by piggybacking on the views of others: that's what *affiliation* is. Consequently, the view seems hard-pressed to explain how, with uses of slurs, speakers are able to form and initiate the bigoted ideologies and groups that define hierarchical social structures. The affiliationist framework also depends crucially upon the presence of a neutral counterpart for every slur and differences of opinion about the merits of the group, ruling out scenarios on which even targets regard themselves as inferior. In §3 I suggested reasons to doubt both[15].

Can affiliationism explain how terms come to be slurs without appealing to a mechanism like contempt crystallization? Nunberg would likely claim that expressions come to be slurs simply because they are terms of choice by racists and other bigots. That is, he'd offer his buck-stopper motto: "Racists don't use slurs because they're derogative; slurs are derogative because they're the words that racists use." A fuller assessment of this fundamental feature of Nunberg's account must be reserved for another occasion.

14 Affiliationism merits extensive discussion I lack the space to address here. I offer a fuller discussion of this sophisticated theory in Jeshion (2016c).

15 Affiliationism also depends crucially on the social dominance of politeness, so that the slur is, in every context, the marked expression. I challenge it in Jeshion (2016c).

REFERENCES

Anderson, L., and Lepore, E. (2013), "Slurring Words", *Nous,* 47, pp. 25-48;
Banse, R., and Scherer, K. (1996), "Acoustic Profiles in Vocal Emotion Expression", *Journal of Personality and Social Psychology*, Vol. 70, No. 3, pp. 614-636;
Bolinger, R. (2015), "The Pragmatics of Slurs", *Nous*, Vol. 49, No. 1, pp. 1-24;
Camp, E. (2013), "Slurring Perspectives", *Analytic Philosophy*, 54, pp. 330-349;
Camp, E. (2016), "A Dual Act Analysis of Slurs", forthcoming In Sosa, D. (ed.), *Bad Words*, Oxford University Press, Oxford;
Copp, D. (2001), "Realist Expressivism: A Neglected Option for Moral Realism", *Social Philosophy and Policy*, 18, pp. 1-43;
Ekman, P., and Heider, K. G. (1988), "The Universality of Contempt Expression; A Replication", *Motivation and Emotion,* 12, pp. 303-308;
Hatfield, E., Caciopp, J., and Rapson, R. (1993), "Emotional Contagion", *Current Directions in Psychological Sciences*, 2, pp. 96-99;
Hatfield, E., Caciopp, J., and Rapson, R. (1994), *Emotional Contagion,* Cambridge University Press, New York;
Hom, C. (2008), "The Semantics of Racial Epithets", *Journal of Philosophy*, Vol. 105, No. 8, pp. 416-440;
Hom, C., and May, R. (2016), "Pejoratives as Fiction", forthcoming In Sosa, D. (ed.), *Bad Words*, Oxford University Press, Oxford;
Jeshion, R. (2013), "Expressivism and the Offensiveness of Slurs", *Philosophical Perspectives*, 27, pp. 231-259;
Jeshion, R. (2016a), "Slurs, Dehumanization, and the Expression of Contempt", forthcoming In Sosa, D. (ed.), *Bad Words*, Oxford University Press, Oxford;
Jeshion, R. (2016b), "The Truth about Slurs", manuscript;
Jeshion, R. (2016c), "The Social Dimension of Slurs", manuscript;
Matsumoto, D., and Ekman, P. (2004), "The Relationship Among Expressions, Labels, and Descriptions of Contempt", *Journal of Personality and Social Psychology*, Vol. 87, No. 4, pp. 529-540;
Nunberg, G. (2017), "The Social Life of Slurs", forthcoming In Fogal, D., Harris, D., and Moss, M. (eds.), *New Work on Speech Acts*, Oxford University Press, Oxford;
Potts, C. (2005), *The Logic of Conventional Implicatures,* Oxford University Press, Oxford;
Richard, M. (2008), *When Truth Gives Out,* Oxford University Press, Oxford;
Saka, P. (2007), *How to Think About Meaning,* Springer, Berlin;
Tirrell, L. (2012), "Genocidal Language Games", In Maitra, I., and McGowan, M. K. (eds.), *Speech and Harm: Controversies Over Free Speech,* Oxford University Press, Oxford, pp. 174-221.

FEDERICO CELLA
Vita-Salute San Raffaele University
f.cella@outlook.com

SLURS: SEMANTIC CONTENT, EXPRESSIVE CONTENT AND SOCIAL GENERICS

abstract

Slurs are offensive expressions targeting individuals on the basis of their membership to certain social groups. Some authors have argued that the offensiveness of slurs is related to the semantic encoding of stereotypes in their meaning. As noticed by Robin Jeshion (2011, 2013a, 2013b), the stereotypical strategies do not seem to provide a satisfactory analysis of slurs' functional traits. Herein, I propose to modify her view by making a distinction between two offensive dimensions of slurs: a negative expressive component encoded in the semantic content and directed toward a certain group of individuals, and the social generics related to that group conveyed as conversational implicatures.

keywords

slurs, social generics, stereotypes, Jeshion, expressive content

Phenomenology and Mind, n. 11 - 2016, pp. 140-148
DOI: 10.13128/Phe_Mi-20114
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Slurs are offensive expressions usually understood to communicate contempt and hatred toward individuals and groups of individuals on the basis of their presumed racial membership ('nigger'), geographical origin ('redneck'), sexual orientation ('faggot') or other features[1]. In recent years, these expressions have become an issue of increasing interest in philosophy of language, since they possess particular linguistic properties[2].

Different authors (Camp 2011; Croom 2011; Dummett 1973; Hom 2008; Tirrell 1999) have argued that the offensiveness of slurs is related to the semantic encoding of stereotypes in their meaning. I will interpret here such encoded stereotypes as *social generics*. This interpretation is motivated by the power of generics to promote *social essentialism* (Rhodes *et al.* 2012) and to characterize their targets in a way which is hard to challenge (Leslie 2007, 2008, 2012, forthcoming; Langton *et al.* 2012; Haslanger 2014). Associating stereotypes to the semantic content of slurs presents some explanatory advantages; nevertheless, the stereotypical strategies do not seem to provide a satisfactory analysis of all the functional traits of slurs.

The difficulties of the stereotypical semantic strategies can be avoided: my proposal is to characterize the semantic content of slurs as analytically decomposable in a *negative attitudinal non-truth-conditional component* directed toward a group of individuals referred to by a descriptive non-stereotypical component. Recently, Robin Jeshion (2011, 2013a, 2013b) has proposed an analysis that takes this direction. Her strategy regards slurs' content as decomposable into several semantic components, treating the activation of stereotypes as a possible contextual association with the use of these expressions – i.e., as *perlocutionary effects* on bystanders. I would suggest modifying Jeshion's theory by distinguishing between two offensive dimensions of slurs: 1) a negative expressive component encoded in the semantic content and directed toward a certain group of individuals, and 2) the association of social generics related to that group as *conversational implicatures*.

1 Some other derogatory expressions like this type will appear in my work, but I need to point out that these terms will be *mentioned* (just for didactic purposes), and not *used* in the course of the discussion.

2 Among the most important features, slurs possess a neutral counterpart (e.g., 'Chink' and 'Chinese'), and the 'derogatory autonomy', i.e. the capacity to derogate independently from the intention of any of its particular speakers. Furthermore, their offensiveness varies over time, and the targets of a slur can use the same term in an appropriate and non-offensive way to signal solidarity between members of the same community (Hom 2008: 426, 428). Other properties, relevant to this topic, will be discussed in the following sections.

2. Slurs and social generics

In order to explain slurs' offensiveness and functional traits, some authors have proposed several treatment strategies according to which the meaning of slurs is related to the semantic encoding of stereotypes, while considering the codification of stereotypes as an inferential association (Dummett 1973; Tirrell 1999), or as a complex descriptive condition (Hom 2008), or as part of a two-dimensional semantics (Camp 2011; Croom 2011). The intuition of what Robin Jeshion calls a "*stereotype semantics of slurs*" is justified at least by five considerations (Jeshion 2013a: 314-315):

(i). The use of slurs is widely spread in communication, as well as the stereotypes related to the groups themselves. Utterances containing slurs easily activate stereotypes in the minds of bystanders, who are likely to infer that the speaker endorses the same existing social stereotypes.
(ii). Slurs and stereotypes seem to present some common features: unlike pejoratives in general (e.g., 'idiot'), they both refer to social categories beyond the individuals to whom they are directed, and often in a derogatory way.
(iii). An appeal to the semantic encoding of stereotypes can explain why slurs are surrounded by a stronger taboo status in society than other pejoratives.
(iv). Assuming that slurring terms semantically encode stereotypes permits to account for slurs' 'derogatory variation'[3]: the offensiveness of different slurs will vary depending on the different offensiveness of the stereotypes encoded.
(v). If slurs semantically encode stereotypes, then it is possible to explain the use of slurring terms whose extension can be restricted only to some in-group members, or expanded to include even out-group members, since it would be sufficient for the individuals targeted to be conforming to the stereotypes a slur ascribes.

These analyses can be reinterpreted in a more explanatory way, considering as part of the semantic content of slurs a particular subset of stereotypes, sharing the syntactic form of the linguistic category of generics.
Generics, which are generalizations that omit quantifiers (like 'some', 'most' or 'all'), behave in a different way, compared to ordinary quantified statements, as they express claims about the "essence" of a particular category (instead of its particular members). The mere absence of the ascribed property does not undermine the content of the generics; therefore, social generics characterize entire target groups in a way which is difficult to challenge (Leslie 2007, 2008, 2012, forthcoming; Langton *et al.* 2012; Haslanger 2014).
Consider the following statements:

(1) All Italians are mobsters
(2) All Muslims are terrorists
and
(3) Italians are mobsters
(4) Muslims are terrorists.

In order to express a generalization about two different groups, a speaker can choose to utter either (1) or (2), focusing on the target groups' members – otherwise she could utter either (3)

3 As highlighted by Hom, "Some epithets are more insulting than others. While epithets like 'nigger' are extremely derogatory towards African-Americans, epithets like 'limey' are much less derogatory towards the English" (2008: 426).

or (4), focusing on the communities themselves. While a single counterexample is sufficient to render the generalization expressed by (1) and (2) false, (3) and (4) cannot be undermined even if the vast majority of the individuals who belong to the relevant target group lack the ascribed property. It does not matter, for example, if the majority of Muslims are not terrorists. With an utterance of (4), the speaker seems to link *being terrorists* with a supposed Muslims' "essence"; and, if we try to challenge the generic with different counterexamples, the proponent of (4) can always affirm that although many Muslims are not terrorists, it does not mean that property is not part of their essence; in fact they would exhibit that property if they were under certain circumstances (because they are "disposed" to possess that property). Recent experimental studies (e.g., Rhodes *et al.* 2012) show that hearing generic language about a social group seems to lead (especially children) to think in an essential and stereotyped way that particular social "categories" mark distinct kinds of people. As a result, social generics – promoting *social essentialism*[4], i.e. the belief that the members of a social kind share stereotypical properties and dispositions due to their supposed "essence" – are difficult to "contrast" because they can characterize entire target groups while allowing for many exceptions.
I believe that the distinctive features of social generics can be usefully linked to the offensiveness of slurs: in addition to being motivated by the five above-mentioned considerations (i)-(v), the intuition that slurs semantically encode stereotypes as social generics can account for two further points:

(vi). If slurs semantically encode social generics, then it is possible to explain the capacity of these terms to characterize and derogate whole target groups in a way which is difficult to "contest".
(vii). Since generics promote *social essentialism*, assuming that slurs semantically encode social generics, we may explain the capacity of slurs to bolster social oppression toward their target groups.

3. Stereotypes as perlocutionary effects

Although this proposal involves some advantages, stereotypical semantic strategies do not seem to provide a satisfactory analysis of all the functional traits of slurs. In particular, as Jeshion highlights (2011, 2013a, 2013b), a stereotype semantics of slurs must deal with several problems.
First: the knowledge of the stereotypes regarding a particular target group does not seem to be a necessary condition to understand a slur. In fact, a speaker may choose to employ a slur in order to express contempt toward one or more individuals on the basis of their membership to a relevant target group, while consistently disavow the approval of any stereotypical feature typically attributed to the same group (Jeshion 2013a: 320, 322). Second: stereotypes cannot be considered as the only source of offensiveness of these expressions, since there are stereotypical beliefs culturally perceived as neutral or positive (Jeshion 2013a: 323 and Jeshion 2013b: 245-246). Third: slurs and particular occurrences of their relative neutral counterparts seem to present the same semantic properties.
Consider the following statements:

(5) John is a nigger
(6) John is *black*^{C}
(7) John is a fucking black.

4 The term is due to Rhodes *et al.* (2012).

In (6) the "C" superscript indicates stress intonation of contempt (Jeshion 2013a: 321). It is undeniable that (6) and (7) express an offensive content toward an individual, John, on the basis of his group membership, as much as (5). However, it is not plausible to maintain that these occurrences of neutral counterparts involve the semantic encoding of stereotypes. If this observation is correct, and if the neutral counterparts marked negatively exhibit the same semantic properties of slurs, then this observation represents a serious concern for any stereotype semantics of slurs. If we hypothesize that stereotypes are semantically encoded in the literal meaning of slurs, then we should say the same for their neutral counterparts – and that seems highly implausible (Jeshion 2013b: 246).

These remarks seem to justify the hypothesis according to which slurring terms could be considered as expressions that linguistically signal the encoding of an expressive negative component directed toward the group denoted by a non-stereotypical descriptive component. Indeed, slurs seem to present an interesting combination of properties generally attributed both to the expressive and descriptive dimension of meaning[5]. On the other hand, the activation of stereotypes could be treated as a contextual association with the use of slurs. Jeshion (2011, 2013a, 2013b) has recently proposed a treatment strategy of slurs that takes this direction. According to her proposal, slurs present a tripartite semantic structure – which is decomposable into a *truth-conditional component*, a *negative attitudinal component*, and an *identifying component*.

The encoding of the truth-conditional component explains why a slurring term conventionally refers to the target group denoted by its neutral counterpart (Jeshion 2013b: 240).

The second component of the semantic content of slurs is attitudinal (non-truth-conditional): a slurring term, unlike its neutral counterpart, is conventionally used to express contempt for the members of a target group because of their membership to that group. Following the characterization of the expressive dimension proposed by Kaplan (1999)[6], Jeshion gives this component the following *rule of use*:

> for a group slurring term S with a neutral counterpart NC that references a group G, S is used to express contempt for members of G on account of their being in G or on account of their possessing a G-defining property g (Jeshion 2013b: 240-241).

The third component of the semantic content of slurs is the identifying one: according to Jeshion, the offensiveness of slurring terms is not limited to the expression of an attitude of contempt toward one or more individuals on the basis of their membership to a particular group. For example, when someone calls someone else 'faggot', the homophobic speaker identifies a property that believes is possessed by that group of people as a defining feature of theirs – possibly the only one. This component

5 *Expressives* are terms conventionally used by speakers in order to highlight their own emotional states. This category includes expressions such as pejoratives, pure expressives (e.g., 'Oops' or 'Ouch'), and honorifics (e.g., 'Madam' or 'Sir'). Contrary to *descriptives* (i.e., emotionally neutral terms used to ascribe features or properties such as 'black', 'popular', and 'small'), expressives do not contribute to the truth-conditions of the utterances in which they occur. Expressives seem to play a key role in communication: they allow speakers to effectively achieve complex social results, and to express their emotional states and perspectives in a more functional way than descriptives (Jay 2009: 153).

6 Among the first to engage in a systematic analysis of expressives, Kaplan (1999) has claimed that the semantic treatment of an expressive requires the identification of its correct conventional use. According to his view, an expression is descriptively correct if what it describes is the case, while expressively correct if the emotional state or attitude displayed by the speaker is the case.

> is used to dictate how others ought to treat, regard, think of, and respond to its target [...]. Notice that the identifying component is dependent upon the expressive component because [...] it follows from what it is to find someone contemptible on the basis of being gay that one takes that person's sexual orientation as the most or among the most central aspects of that person's identity. This dependence is, at heart, moral-psychological, but is manifest in the semantics. A speaker who expresses contempt toward her target for being G *thereby* also expresses and implicitly represents G as fundamental to her target's identity as a person (Jeshion 2013b: 242).

Jeshion (2013b: 242-243) maintains that the offensiveness of this component is closely associated to the offensiveness of the attitudinal component. Indeed, according to her view, what makes the use of a slur dehumanizing to its target group is the conjunction of these two components into a single utterance. However, the identifying component explains the particular offensiveness of slurs to target individuals who do not consider the identity of the target group as a defining feature of their person, or do not want someone else to define their social identity from the outside[7].

According to Jeshion, there would be also pragmatic sources of the offensiveness of slurs – which do not depend on the conventional meaning or on the rules of use of these terms. In particular, according to her treatment strategy, the use of slurs can contextually activate the stereotypes related to relevant target groups – in Austin's terms – as *perlocutionary effects* on bystanders (Jeshion 2013b: 245-247). In other words, the utterance of a slur by the speaker can change the beliefs, emotions and actions of the hearers in a socially and psychologically predictable way (although this may happen unintentionally or even not at all), in particular activating stereotypes and, among other possible effects, compounding the social oppression toward the relevant target groups.

4. Social generics as conversational implicatures

The treatment strategy proposed by Jeshion presents several advantages: considering the content of slurs as decomposable into a truth-conditional neutral component and an offensive attitudinal component, this strategy helps to explain why the offensiveness of such expressions persists across negation, denial, conditional and other complex constructions – since the expressive component cannot be "harnessed" at the truth-conditional level. Moreover, treating stereotypes as possible contextual associations to the use of slurs – i.e. as perlocutionary effects on bystanders – Jeshion's analysis avoids the problems presented by stereotypical semantic strategies[8]. However, this proposal presents some difficulties too.

First: in order to explain the functional traits of slurs, the semantic encoding of the identifying component does not seem necessary to me. If someone employs a slur, the attitudinal component will be directed toward the ascription to the target group, and since the speaker is already categorizing the membership to a group to be the – or one of the – defining characteristics of the identity of their target, this additional component is not necessary to characterize the semantic content of these expressions. In certain contexts, even the ascription of the neutral counterpart, and in general of any predicate – like 'student' –, may be offensive to those who do not consider this feature as their defining characteristic or do not want their identity to be defined by others. Moreover, since Jeshion maintains that slurs

7 Actually, as we shall see in the next section, this also seems to apply for terms that are not slurs, like neutral counterparts.

8 Jeshion's theory is far more complex than displayed here and it presents more actual advantages; nevertheless, what is shown is sufficient for my purposes.

and relative neutral counterparts marked negatively have the same semantic properties, it should be argued that the contempt intonation allows to semantically associate to neutral counterparts the identifying component, in addition to the expressive one – which does not seem particularly plausible. Furthermore, the addition of this component also seems problematic: maintaining that the semantic content of slurs prescribes how their target groups ought to be treated, this seems to commit Jeshion herself to support a stereotype semantics of slurs, as if these expressions included essentializing normative generics. Briefly, this component should also be considered as a possible contextual association with the use of slurs, instead of a semantic component of these expressions[9].

Second: although the proposal to consider both the stereotype activation and the compounding of social oppression as pragmatic sources of the offensiveness of slurs is acceptable, treating these phenomena as possible perlocutionary effects associated to the use of slurs seems overly reductive. The limits of such a proposal are particularly clear when it is related to three points: that solution cannot explain 1) *why* the use of slurs (or neutral counterparts negatively marked) can increase social oppression toward their targets and 2) *why* the utterance of slurs (or neutral counterparts marked) in certain contexts activate some of the existing social stereotypes but not others, and 3) *how* such an activation can be understood by the speaker through the exploitation of the extra-semantic characteristics of a conversation.

I do believe that these difficulties can be avoided by modifying Jeshion's theory, distinguishing between two offensive dimensions of slurs: 1) the semantic encoding of a negative expressive component directed toward a group of individuals referred to by a descriptive component and 2) the association of the social generics related to the target group as Gricean *conversational implicatures* (Grice 1975).

This characterization is compatible with Jeshion's observation that slurs and neutral counterparts negatively marked seem to have the same semantic properties: in fact, the association of the contemptuous intonation to the neutral counterpart of a slur does not seem to require the encoding of an identifying component in addition to the descriptive one. If slurs and negatively marked neutral counterparts have the same semantic properties, then the encoding of the identifying component can be rejected – avoiding the problem of explaining the encoding of the prescriptive practices assumed by Jeshion. I would like to consider this modification both as non-problematic and advantageous: indeed, this variation allows solving, in a simple and effective way, one of the problems associated with Jeshion's proposal.

Moreover, this characterization permits to avoid another difficulty in Jeshion's analysis.

The characteristics of conversational implicatures seem particularly suited to explain how a speaker can use a slur to convey an additional stereotypical content.

First: conversational implicatures can be *canceled* (explicitly or contextually); this characteristic could explain why a speaker may use a slur to express contempt for one or more individuals on the basis of their mere membership to a target group without being aware of any particular stereotypical characteristic related to the same group, or deny the approval of these features without contradiction (while still expressing an offensive content toward the target group, due to the encoding of a negative expressive component). Conversational implicatures explain this possibility since they are not part of the semantic content of the

9 A clarification: Jeshion maintains that the identifying component, in fact, "should not be conflated with any notions of metaphysical identity or essentialism [...]. Rather, they [slurring terms] express that the target's group membership is the, or among the, most central characteristic(s) for classifying what the target is" (Jeshion 2013b: 242). However, my observation avoids this problem because, as highlighted by Leslie (2008, forthcoming), generic claims are purely psychological, not metaphysical: "human perception is the important factor here" (2008: 34).

utterances they are associated with.
Second: conversational implicatures are *not detachable*; using a different but propositionally equivalent statement – for example,

(8) Vincent is a Wop
and
(9) Vincent is *Italian*^{C}

in the same communicative context generates the same implicature. This could explain why, for example, answering the question

(10) Why do you think Vincent should not be hired?

with something like (8) or (9) would convey in the same context – and responding to the same question – the same stereotypes.
Third: the propositional content of a conversational implicature presents independent truth-conditions that do not contribute to the truth-conditional meaning of what is explicitly said; (8) and (9) can be true or false regardless to the truth-value of the conversationally implicated content. This could explain why a slur can be used against an individual the speaker does not consider as belonging to a specific target group, in order to ascribe her a stereotypical property that the speaker believes to be true.
Fourth: conversational implicatures can be *calculated* – i.e., it is possible to reconstruct a series of inferential steps that guided the hearers to derive the implicature and which may be expected by the speaker.
Finally, this characterization offers a possible explanation to the reason why the use of slurs (or of the negatively marked neutral counterparts) can bolster social oppression toward their targets. Indeed, considering the social stereotypes conveyed as social generics, it is possible to explain why the offensiveness communicated by slurs is so difficult to challenge. Summing up, we can say that, given that generics promote social essentialism, i.e. the belief that the members of a social kind share stereotypical properties and dispositions due to their supposed "essence", this additional source of offensiveness can easily explain the capacity of slurs to increase social oppression toward their targets.

5. Concluding remarks

In this work, I have suggested considering the semantic content of slurs as decomposable into an offensive expressive component directed toward a target group, denoted by a descriptive component. This semantic structure seems to be confirmed, as noted by Jeshion, by neutral counterparts (which are truth-conditionally equivalent to the descriptive content of slurs) marked with an external expressive component, which present the same semantic properties of slurs. Moreover, I believe that slurs (and derogatory occurrences of the neutral counterparts) can be offensive in an additional way, linked to their context of utterance – i.e. their use may convey social generics related to the relevant target group as conversational implicatures. Since generic language promotes social essentialism, this additional source of offensiveness can explain the capacity of slurs to increase social oppression toward the target groups they are directed to.
This characterization helps avoid the main problems both in stereotypical semantic strategies and in the treatment strategy proposed by Jeshion. However, this proposal of mine needs the resolution of two problems, which would require further extensive discussion far beyond the goal of the present work. Slurs, and pejoratives in general, can be used ironically in a friendly (but non-appropriate) way among out-group individuals in order to increase their

social intimacy: in these contexts, the use of slurs is fundamentally based on the ascription of stereotypical features contextually salient. Therefore, the association of social generics is very significant in the context of utterance and does not seem to be cancelable by the speaker himself without contradiction. If this observation were correct, it would raise a possible objection against the possibility to consider the association of social generics as conversational implicatures. Finally, my proposal shares one limit also present in Jeshion's treatment strategy: as a matter of fact, both proposals are applied to slurs whose descriptive content only denotes the group referred to by their neutral counterparts and only to those occurring as nouns. This aspect does exclude the analysis of those expressions that appear to encode an additional descriptive content, such as "Beaner" or "Christ killer", or other slurs used with different grammatical functions. The slurring terms to which this proposal of mine might be applied are the "central" expressions of this lexical category, but an adequate theoretical analysis should consider these terms as widely as possible. The view outlined above requires a development toward this direction.

REFERENCES

Camp, E. (2011), "Slurs, Semantics, and Stereotypes", *Syracuse Philosophy Annual Workshop and Network 2011*, Syracuse University, Syracuse, New York;
Croom, A. M. (2011), "Slurs", *Language Sciences*, 33(3), pp. 343-358;
Dummett, M. (1973), *Frege: Philosophy of Language*, Harper & Row, New York;
Grice, H. P. (1975), "Logic and conversation", In Cole, P., and Morgan, J. L. (eds.), *Syntax and Semantics 3: Speech Acts*, Academic Press, New York, pp. 41-58;
Haslanger, S. (2014), "The normal, the natural and the good: Generics and ideology", *Politica & Società*, 3(3), pp. 365-392;
Hom, C. (2008), "The semantics of racial epithets", *The Journal of Philosophy*, 105, pp. 416-440;
Jeshion, R. (2011), "Dehumanizing slurs", *Society for Exact Philosophy meeting,* Winnipeg, Manitoba;
Jay, T. (2009), "The utility and ubiquity of taboo words", *Perspectives on Psychological Sciences*, 4(2), pp. 153-161;
Jeshion, R. (2013a), "Slurs and stereotypes", *Analytic Philosophy*, 54(3), pp. 314-329;
Jeshion, R. (2013b), "Expressivism and the offensiveness of slurs", *Philosophical Perspectives*, 27(1), pp. 231-259;
Kaplan, D. (1999), "The meaning of Ouch and Oops", *Howison Lecture*, UC Berkeley;
Langton, R., Haslanger, S., and Anderson, L. (2012), "Language and Race", In Russell, G., and Graff Fara, D. (eds.), *The Routledge Companion to the Philosophy of Language*, Routledge, London, pp. 753-767;
Leslie, S. J. (2007), "Generics and the structure of the mind", *Philosophical Perspectives*, 21(1), pp. 375-403;
Leslie, S. J. (2008), "Generics: Cognition and acquisition", *Philosophical Review*, 117(1), pp. 1-47;
Leslie, S. J. (2012), "Generics", In Russell, G., and Graff Fara, D. (eds.), *The Routledge Companion to the Philosophy of Language*, Routledge, London, pp. 355-366;
Leslie, S. J. (forthcoming), "The original sin of cognition: Fear, prejudice, and generalization", *The Journal of Philosophy*;
Rhodes, M., Leslie, S. J., and Tworek, C. M. (2012), "Cultural transmission of social essentialism", *Proceedings of the National Academy of Sciences*, 109(34), pp. 13526-13531;
Tirrell, L. (1999), "Derogatory Terms. Racism, Sexism, and the Inferential Role Theory of Meaning", In Hendricks, C., and Kelly, O. (eds.), *Language and Liberation: Feminism, Philosophy and Language*, SUNY Press, Albany, pp. 41-79.

MIHAELA POPA-WYATT
LOGOS, University of Barcelona
popa.michaela@gmail.com

NOT ALL SLURS ARE EQUAL

abstract

Slurs are typically defined as conveying contempt based on group-membership. However, here I argue that they are not a unitary group. First, I describe two dimensions of variation among derogatives: how targets are identified, and how offensive the term is. This supports the typical definition of slurs as opposed to other derogatives. I then highlight problems with this definition, mainly caused by variable offence across slur words. In the process I discuss how major theories of slurs can account for variable offence, and conclude that contempt based on group-membership doesn't cover all the data. I finish by noting that the most offensive slurs are those that target oppressed groups. I claim it is oppression that underpins most offence, and that beyond this offensive property, some slurs are actively used to oppress.

keywords

slurs, derogatives, variable offence, oppressive slurs

Phenomenology and Mind, n. 11 - 2016, pp. 150-156
DOI: 10.13128/Phe_Mi-20115
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Slurring is a kind of hate speech. Examples most commonly cited in the literature are those based on race and ethnicity (“nigger”, “kike”, “chink”, “spic”, “paki”), gender (“bitch”), and sexual orientation (“dyke”, “faggot”, “tranny”). Slurs, although pejorative in a broad sense, are commonly deemed to be different from pejorative words (“jerk”, “dickhead”, “asshole”, “wimp”). In addition, there are words that are clearly derogatory (“fatso”, “wino”, “druggie”, “slut”, “commie”), but where there is no universal agreement about the category (pejorative or slur) to which they belong. Finally, there are some words with multiple uses, all likely to cause offence, one of which is a slur. The most obvious example is “cunt”, which can be used either as a swearword, as a vulgar noun, as a pejorative, or as a slur.
Are slurs different from other derogatives? If slurs are special, what makes them special? In this short paper, I will try to answer this question. I will come to the conclusion that not all words referred to as slurs are special, but that some are. Notably, these are the slurs that are most hotly discussed.

2. Derogatives

Derogative expressions are evaluatives, they express negative attitudes towards the target. Examples include [1] to [6]:

[1] “Faggot!”
[2] “He’s a Jew^{C}!” (C marks contemptuous intonation).
[3] “Filthy fucking Jew!”
[4] “I hate gays.”
[5] “You dickhead!”
[6] “Hey fatso!”

All the above are expressions of derogation. What makes them different from one another? There are two dimensions of variation between these examples that will be important for my argument.
The first dimension is in how the target is identified. In (1) the target is identified by membership of a group. This property of derogation based on group-membership is characteristic of the word itself, not a property of its use. To see the difference note that in (2) contempt is expressed, and that contempt clearly arises from group-membership, but

the expression of contempt is by tone. The word "Jew" is perfectly capable of a neutral use.[1]
A similar effect to (2) can be achieved by prefacing with expletives as in (3), or by explicitly reporting what attitude one has towards the target, say, hate in (4). All of examples (1)-(4) concern groups. What they have in common is that they express a negative attitude not just towards an individual, but to a group. (4) does this explicitly, but (1)-(3) do this by implication: if I hate you because you are a Jew, by default I can be expected to hate all Jewish people.
Note however that though similar slurring effects arise from using a neutral counterpart with contemptuous intonation, or accompanied by expletives as in (2)-(3), they differ substantially from a slur term as in (1) in that only the latter has contempt conventionally attached to the word as based on the target's group-membership. For example, the anti-Semite saying "Kike" shows contempt for Jews precisely because they are Jews. Contempt and group-classification are bound together in a slur term so that one cannot do without the other. Not so in (2)-(3), where contempt is disjoint from the neutral counterpart, deriving instead from external sources to do with the *speaker's own expression of contempt*. Remove the contemptuous intonation or the expletives, and the offence falls out. The offence would still persist even if we were to remove the expletives from "goddamn' fuckin' fags".
On the other hand, (5) targets a particular individual and expresses contempt based on personal qualities of that individual. There is no reasonable sense in which we could say that (5) derogates the group of all dickheads, in the way that (1) derogates all the gay people.
Finally, (6) sits in the middle ground between (5) and (1)-(4). If someone is overweight, calling them "fatso" is certainly targeting them as an individual, but it is again by implicitly expressing a negative attitude towards all people who are overweight. In consequence, while there is a coherent movement against "body-shaming", there is no such movement for the protection of "dickheads". There are quite a number of other derogatives that fall in this middle ground. These include "wino", "druggie", "commie", "fatso", "psycho". Like pejoratives targeted at individuals, they identify the targeted individual on the basis of specific properties that s/he has. But like slurs, they express contempt not only about the particular individual but also about other people who have similar features, and so may be identified as part of a group.
The second dimension of variation is in the intensity of the effects achieved. Most notably, different words cause different degrees of offence, and in different patterns. Calling someone a "dickhead" is likely to offend them, and probably any close friends. Calling someone a "nigger" is considered to be much more offensive, and is offensive to any non-racist who hears it. There are also strong differences within the group of words that are group based. For example, as noted by Jeshion (2013), "Nigger" is widely considered to be more offensive than "Spook" and "Chink", which are in turn more offensive than "Limey" and "Yankee". So variation in offence is not just between slurs and other pejoratives, but across slur terms themselves.
So, we have two types of differences we can readily identify between derogative words. First, how the target is identified. Second, how offensive the word is.
On this basis it seems reasonable to follow authors like Jeshion, and say that a slur is defined as a word that is used to express contempt on the basis of group-membership. On that account we would have to allow that the class of derogatives based on properties that many individuals share are at least slur-like. This seems on the face of it a satisfactory explanation of what makes slurs different from individual pejoratives.
But this does not explain offence variation across slur terms. This is the knottier problem. There are variety of ways of dealing with it. I consider these now.

1 Given its pedigreed history, the neutral term "Jew" has acquired a negative connotation. Similarly, "Indian".

3. Accounting for Variable Derogation

There are several accounts of variable derogation. The ones I will consider are semantic, pragmatic, and prohibitionist.

On a semantic account the variation in offence is a product of the valence of a stereotype encoded by a slur word. This is the sort of position that we would take if we were to follow Chris Hom in his original 2008 paper on slurs. In that paper he argues that a slur semantically encodes a stereotype—i.e. that it predicates a number of negative properties together with a set of deontic prescriptions that are externally derived from practices and institutions of oppression. Thus, the term "Chink" predicates that the target ought to be discriminated for a number of negative stereotypical properties all because of being Chinese. This accounts nicely for variable offence across slur terms. Thus the argument would be that "Nigger" contains a more negative stereotype than "Chink", "Spook", or "Limey". In particular, the pejorative force of a slur word varies given differences in the valence of the semantic content—i.e. the amount of negative properties (and consequently the number of deontic prescription) that a racist institution conventionally encodes into a slur.

This makes sense as far as it goes, but it struggles to account for all the facts. This is because there is a second kind of variation that does not have to do with the words themselves, but with their use. For example, in-group uses may cause no offence. If a gay man refers to a gay friend as a "faggot", and the phrase is clearly uttered without contempt, it may well not be considered offensive. The same is true of African-Americans using the term "Nigger" to refer to one another. But if a white person uses the term to a African-American friend it is much more likely to be considered offensive. This causes problems for a strict semanticist, since if the offence is really encoded in the word it should be the same offence regardless of who uses the word.

This is strongly suggestive that some portion of the offence is carried by pragmatics. One can imagine that an account given in terms of conversational implicature might account for the lack of offence of in-group reclaimed uses. However, it is perfectly possible for a gay speaker to use [1] to derogate another gay target, so this reasoning isn't sound. The fact that a slur is used by in-group members doesn't automatically make it inoffensive. So, even when the causal link between the slur word and the stereotype (along with associated discriminatory practices) is severed on account of the speaker's group-membership, there is still room for the speaker to use the word to express their own contempt. Another route is to claim that there are in fact two words, and this is supported, for example, by the fact that there are two spellings of some words: "Nigger" and "Nigga" with the second being used exclusively to refer in a friendly, in-group context. However, the fact that the second word is not open to use by out-group members suggests that the term is not, in reality a separate word.

What we need to explain is how the word acquires a new convention of use that serves a different purpose for the in-group community. What would that other purpose be? Bianchi (2014) suggests this may be ironic self-deprecation and jokingly self-mockery so as to take the sting out of the ridicule and disdain characteristic from out-group members. This is achieved, she argues, through echoing the derogatory uses by bigots in ways that make manifest their dissociation from the offensive contents. This is clearly a fairly pragmatic matter that is sensitive enough to speaker's intentions such that it leaves room for in-group members to express their own contempt, and hence offend in no dissimilar ways than one would offend with an individual pejorative.

There are other semantic theories of slurring offence. The most notable of these is the work of Jeshion, who argues for an expressivist semantics. The critical part of this view is that it encodes the expression of contempt semantically. This expressivist view has been attacked for not being able to account for variable derogation across words. The argument against is as follows. If the semantic expression of contempt is uniform across all slurs in that a common

core attitude is expressed, then Jeshion's semantics cannot explain variable offence as being caused by variable contempt. Critics claim that this makes it impossible for expressivism to explain variable offence, and thus reject it (Hom 2008).

There are two ways out of this for the expressivist. First, it could be maintained that different words encode varying degrees of contempt, and that it is the degree of contempt expressed that causes the variation in contempt. Jeshion does not propose this, but it seems *prima facie* that there is nothing wrong with this approach. Certainly it seems intuitive to talk about varying degrees of contempt. People regularly use expressions such as "utter contempt", "complete contempt", and "beneath contempt". This suggests that contempt may at least be a property capable of variation, whether or not that can be semantically encoded in words that signal it.

One option for expressivism is to model this variation along similar lines to the variation in the degrees of emotion expressed with ordinary evaluative terms (e.g. "good", "great", "wonderful"). Potts (2007) applies this model to expressive words such as "damn", "hurray". Imagine, for example, someone uttering "Trump is a damn republican". "Damn" signals here a negative shift in the speaker's attitude towards Trump's being a republican. However, if the speaker is a Trump-supporter she might utter "Trump is a damn fine republican" to convey that she has a positive attitude. Thus, what explains the variation is a function shifting the "expressive index" of the conversational context so that an expressive word may be taken to semantically express varying degrees of intensity of the emotion expressed in a context, and how they may differ across contexts. On this model then, using a slur could be treated as updating the common ground with information about the speaker's negative attitudes toward the targeted individual, and toward members of the group as a whole. Thus, different slurs may be conventionally associated with varying degrees of strength of negativity depending on the stereotypes that are drawn upon as the basis for the negative attitudes conventionally expressed.

A second way out, pointed out by Jeshion herself, is through pragmatics. This would provide a different way to handle the issue of variability of offence across uses. Thus, different contexts would modify the contempt expressed semantically. It could be that such a route is complementary to the first, that different words may express variable contempt, and also that this contempt expressed semantically may be further modulated by the context.

However, there is a problem if we ground offence in contempt expression alone. It is that there are many words that can be used to express great contempt, and yet which are not nearly as offensive as slurs. I can call someone the "scum of the earth", but I will not be ostracised by polite society for so doing. This suggests that there really is something else going on than contempt expression, although contempt expression is clearly part of the picture.

This is what leads us to another route, that of prohibitionism. Anderson and Lepore (2013) propose that what makes slurs offensive is that they are prohibited, and that different words are prohibited to different degrees. Prohibitionism neatly accounts for why slurring uses are also offensive, for example, during speech reports. What prohibitionism doesn't neatly account for is why some uses are acceptable, such as the reclaimed uses mentioned above. The argument made is that there are exceptions on the prohibition in certain circumstances, so that the prohibition is lifted for in-group uses. This allows consistency with the data, but at the price of a satisfying explanation.

In fact, the route into explaining why slurs differ so greatly in their offence, while also accounting for their group identifying nature, lies elsewhere.

4. Some slurs are not like others

Consider the slur terms that most deeply offend: "Nigger", "Chink", "Bitch", "Wetback", "Faggot". Now consider this paragraph from the "Five Faces of Oppression" by Iris Young (1990: 42):

> I offer some explication of the concept of oppression as I understand its use by new social movements in the United States since the 1960s. My starting point is reflection on the conditions of the groups said by these movements to be oppressed: among others women, Blacks, Chicanos, Puerto Ricans and other Spanish speaking Americans, American Indians, Jews, lesbians, gay men, Arabs, Asians, old people, working-class people, and the physically and mentally disabled.

Is it coincidence that many of the most offensive slur words are associated with groups we might identify as oppressed? I suggest that it is not. Consider slurs for groups that have been oppressed: "Redskin", "Kike", "Faggot", "Towel-head", "Chink". These number among the most offensive words we have. Then consider the existence of slurs for groups that historically have not been oppressed: "Yankees", "Brits", "Bosch", "Ivan". There are also numerous derogatives for the rich and powerful: "toff", "Hooray Henry", "toffee nose", "upper class twit", "filthy rich". Yet despite conveying considerable contempt these words are, frankly, mildly offensive at best. Neither are they prohibited. Why is that? My suggestion is that it is socially acceptable to slur the powerful, precisely because they are powerful, and it does them no harm.

The evidence thus points to a differentiation between words on the basis of oppression. Slurs are not a unitary group of words on the basis of group membership. There are a variety of slurs, and only some of them are deeply offensive. The common pattern is that words that we prohibit are words that refer to the oppressed. Thus, my claim is that not all slurs are equal. Some slurs refer to oppressed groups, and some do not. The first are broadly considered offensive, and the second are not, or at least to a lesser degree. It is oppression that is the fuel of deep offence.

In one sense this is not yet new; most slurs are written about in the literature precisely because they are words used to target oppressed groups. But by focusing on slurs that target oppressed groups we ignore that many slur terms do not. Highlighting this should change how we delineate between slurs and other derogative terms.

But there is some subtlety required here. First of all, there are many words that do refer to groups we should objectively identify as oppressed, but which, at the current time, are seen by many in polite society as acceptable to use. Thus, take words such as "wrinkly', "old fart", "coffin dodger", "kev", "chav", and so on. These words also target oppressed groups, those who are marginalised or exploited. Yet many who would find racial slurs offensive, do not find these words offensive. This needs to be explained. It may be that, unfortunately, the degree of oppression suffered by these groups is considered by polite society to be acceptable. If this were the case, then the degree of offence is related to the degree to which oppression is seen as being socially unacceptable. This is a complex phenomenon, which will require unpacking.

There is another consequence of this division into words that target oppressed groups, and words that target non-oppressed groups. It is that the former group of words do not just seek to describe oppression, they seek to actively create it and maintain it. Thus, some slurs are part of the mechanism of oppression. They are actively part of the process of oppressing other groups. To see this we need only consider the way that slur terms and hate speech are being used in contemporary political discourse. Specifically, if we consider contemporary events

in politics[2], slurs, and derogatives more generally, are being used as tools for incitement to persuade one group that sees itself as disempowered to gain power by oppressing another. Scapegoating and hate speech are used to silence others by fear, and as ways of emotionally appealing to would-be bigots to join the side of the bigot. For example, Trump supporters have been quoted as referring to Muslims as "Muzzies" and "Mo-slimes". A journalist noting this, who happened to be Jewish, received messages such as "Milbank is an anti-white parasite and a bigoted kike supremacist" (Milbank 2015). This journalist also records attendees at an event assault a Black Lives Matter protester while shouting "Light the motherfucker on fire", and shouting to Latinos "motherfucking tacos—go back to Mexico". This highlights that the purpose of slurs is not one merely of derogation, but one of oppression. Slurs unleash a propensity for violence, create fear in their targets, and make it more permissible for others to give voice to their bigotry to the extent that it is treated as acceptable by political leaders. This ability to incite to action is discussed by Tirrell (2012), who argues that "deeply derogative terms" are words that serve not only to enjoin others to hatred but also to action. In the Rwanda genocide dehumanising slurs have actively served as "action-engendering" and have thus been "part and parcel of genocide, not only an antecedent of it". They motivated Hutu to action by depriving Tutsi of their humanity: by labelling them as "inyenzi" (cockroach) or "inzoka" (snake), this granted permission to hate and a mandate to kill. As Tirrell says (2012: 174), "understanding these speech-acts helps to illuminate the important ways that power is enacted through discourse, how speech-acts can prepare the way for physical and material acts, and how speech generates permissions for actions hitherto uncountenanced."

My closing contention is that we thus need to distinguish between the broader group of slurs, and a smaller group that I will call *oppressive slurs*. Further, I claim that what is centrally interesting about oppressive slurs is how they contribute to achieving oppression. I will close by suggesting that a promising route is via an extension to speech act theory. Words, as pointed by Austin (1962), don't describe the world, they alter it. Exactly how is a matter for enquiry.

REFERENCES

Anderson, L. and Lepore, E. (2013), "Slurring Words", *Noûs* 47, pp. 25-48;
Austin, J. L. (1962), *How to do things with words*, Clarendon Press, Oxford;
Bianchi, C. (2014), "Slurs and Appropriation: An Echoic Account", *Journal of Pragmatics* 66, pp. 35–44;
Hom, C. (2008), "The Semantics of Racial Epithets", *Journal of Philosophy* 105, pp. 416—40;
Jeshion, R. (2013), "Expressivism and the Offensiveness of Slurs", *Philosophical Perspectives*, 27, pp. 231-259;
Milbank, D., December 18, 2015. https://www.washingtonpost.com/opinions/trump-brings-bigots-out-of-hiding/2015/12/18/8a02e4ac-a587-11e5-9c4e-be37f66848bb_story.html
Potts, C. (2007), "The Expressive Dimension", *Theoretical Linguistics* 33(2), pp. 165-97;
Tirrell, L. (2012), "Genocidal Language Games", In Maitra, I., and McGowan, M. K. (eds.) *Speech and Harm: Controversies over Free Speech*, Oxford University Press, Oxford, pp. 174-221;
Young, M. (1990), "Five Faces of Oppression.", In *Justice and the Politics of Difference*, Princeton University Press, Princeton, pp. 39-65.

2 At the time of writing the UK has just voted to leave the EU, primarily to reduce immigration, and there has been an upsurge in incidents of racial hatred. At the same time Donald Trump is running to be president of the USA, and has been criticised for using language that incites hatred of immigrants, Latinos and Muslims.

BIANCA CEPOLLARO
Scuola Normale Superiore (Pisa)/Institut Jean Nicod (Paris)
bianca.cepollaro@gmail.com

BUILDING EVALUATION INTO LANGUAGE[1]

abstract

In this paper I spell out the conditions for a uniform analysis of thick terms and slurs, presented in Cepollaro and Stojanovic (2016). Our claim is that thick terms and slurs convey evaluations via presupposition and represent a device through which language implicitly conveys linguistically encoded evaluations. I introduce the presuppositional account (section 2) and elaborate on the conditions that need to be fulfilled for slurs and thick terms to be analyzed along similar lines (section 3). I discuss the predictions that this approach offers about the issues of reference and extension (section 4). I conclude with some considerations about the role and functions of slurs and thick terms with respect to moral systems (section 5).

keywords

evaluatives, presuppositions, reference, slurs, thick terms

Phenomenology and Mind, n. 11 - 2016, pp. 158-168
DOI: 10.13128/Phe_Mi-20116
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The aim of the present paper is to spell out the conditions for a uniform analysis of thick terms and slurs, that represent a device through which language can convey evaluations in a way that is linguistically encoded and implicit. In Cepollaro and Stojanovic (2016) we presented a presuppositional account of "hybrid evaluatives", a class of terms including slurs and thick terms. Slurs are usually defined in philosophy of language and linguistics as derogatory terms targeting individuals and groups on the basis of their belonging to a certain category. Prototypical examples in English are "chink", "faggot", "kike", "nigger", "wop", etc. Thick terms, on the other hand, have drawn the attention of scholars mainly in metaethics rather than philosophy of language and linguistics. They are often introduced by means of examples (typically: "chaste", "courageous", "generous", "lewd", etc.), rather than through a canonical definition. Roughly, thick terms are expressions that combine descriptive content and evaluative content; unlike slurs, the evaluative component can display different polarities (i.e. it can be positive or negative) and they often lack a single-term counterpart: their descriptive content usually corresponds to a paraphrase. The main claim put forward in Cepollaro and Stojanovic (2016) is that hybrid evaluatives systematically convey evaluations through presuppositions. After briefly introducing the account (section 2), I elaborate on the conditions that need to be fulfilled for slurs and thick terms to be analyzed along similar lines (section 3) and I discuss the predictions that this presuppositional approach can offer about the issues of reference and extension (section 4). In section 5, I conclude by showing how the presuppositional account brings up some general considerations about the role and the functions of hybrid evaluatives with respect to our moral systems[1].

1 I would like to thank the organizers of the San Raffaele Spring School of Philosophy (SRSSP) 2016 on Emotions, Normativity, and Social Life for the great event they put together. Thanks to all the participants for their interesting talks, stimulating discussion and helpful comments. Thanks also to Claudia Bianchi, Robin Jeshion, Uriah Kriegel, Isidora Stojanovic and an anonymous referee for their insightful suggestions. This work was supported by the Labex and Idex grants: anr-10-labx-0087 iec, anr-10- idex-0001-02 psl*, and project mineco ffi2012-37658.

2. A uniform analysis of slurs and thick terms

In Cepollaro and Stojanovic (2016) we analyze slurs and thick terms as belonging to the same kind of expressions that we call "hybrid evaluatives". Our proposal is to treat both of slurs and thick terms as presuppositional triggers. Hybrid evaluatives have a descriptive content at the level of truth conditions. In the case of slurs, it usually amounts to their neutral counterpart ("German" for "boche", "Jewish" for "kike", "Italian" for "wop" and so on). For thick terms we usually lack a single-term counterpart, but we can formulate a paraphrase for the descriptive content. So, for example, the descriptive content of "lewd" is something like "sexually explicit"; the descriptive content of "chaste" is "abstaining from sexual relations", etc. In addition to this descriptive content, hybrid evaluatives also convey an evaluative content, which we analyze in terms of presuppositions: "wop" presupposes something like "Italians are bad for being Italian" and "lewd" presupposes something like "sexually explicit people and things are bad for being sexually explicit", etc.
The main argument in favor of a presuppositional analysis of slurs and thick terms is their projective behavior: the evaluative content survives semantic embedding like presuppositions. Presuppositions are known for projecting out of certain embeddings, such as negation, question, antecedent of a conditional and modal, a.k.a. the S-family (Chierchia and McConnell-Ginet 1990, Simons *et al.* 2010). The presupposition triggered by "stop" amounts to something like (π_1) and survives all the following embeddings:

(1) Madonna stopped smoking.
(2) Did Madonna stopped smoking?
(3) If Madonna stopped smoking, it will be good for her health.
(4) Madonna might have stopped smoking.
π_1. Madonna used to smoke.

Thick terms display the same pattern: all these embedded and non-embedded occurrences of "lewd" presuppose (π_5) (see *i.a.* Gibbard 1992 and Väyrynen 2009, 2013)[2]*:*

(5) Madonna's show is lewd.
(6) Madonna's show isn't lewd.
(7) Is Madonna's show lewd?
(8) If Madonna's show is lewd, I wouldn't like to go to her concert.
(9) Madonna's show might be lewd.
π_5. People and things that are sexually explicit are bad as such.

This projection phenomenon is typically acknowledged also for slurs (see *i.a.* Croom 2011, Richard 2008, Jeshion 2013):
(10) Madonna is a wop.
(11) Is Madonna a wop?
(12) If Madonna is a wop, I wouldn't like to go to her concert
(13) Madonna might be a wop.
π_{10}. Italians are bad in some relevant respect because of being so.
The criterion to establish whether an expression counts as a hybrid evaluative or not is the following:

2 Examples (5)-(13) are from Cepollaro and Stojanovic (2016).

> In addition to a descriptive content, hybrid evaluatives systematically convey an evaluative content that scopes out when embedded under negation, conditionals, modals and questions (Cepollaro and Stojanovic 2016).

One important prediction of the presuppositional account is that when presuppositions like (π_1), (π_5) or (π_{10}) are triggered and are not rejected in a conversation, they become common ground. Imagine that in a conversation someone says: "Bianca? She's a wop". The presuppositional account that I have just sketched predicts that if the participants to such a conversation do not protest, they become somehow complicit in the discriminatory attitude of the person who uttered the slur. The phenomenon is often label as "complicity" in the literature about slurs:

> just hearing (...) *[the slur 'nigger']* can leave one feeling as if they have been made complicit in a morally atrocious act (Croom 2011: 343).
> slurs make recalcitrant hearers feel complicit in the speaker's way of thinking (Camp 2013: 333).
> In some cases, merely overhearing a slur is sufficient for making a non-prejudiced listener feel complicit in a speaker's slurring performance (DiFranco 2014).

Note that in order to account for complicity, it is crucial to establish what is exactly being presupposed; in the literature, two options were put forward, that we can call the "objective" and "subjective" option (see Predelli 2010 and Cepollaro 2015). For "wop" and "lewd", the objective and subjective options would be respectively:

(W_o) Italians are worthy of contempt for being Italian.
(W_s) *The agent believes that* Italians are worthy of contempt for being Italian.
(L_o) People and things that are sexually explicit are bad for being sexually explicit.
(L_s) *The agent believes that* people and things that are sexually explicit are bad for being sexually explicit.

The "subjective" option, proposed in Schlenker (2007) for slurs, has no *direct* way to explain the complicity phenomenon: if the presupposed content is not about targets (Italians, or sexually explicit things), but it is about the beliefs of the speaker, it is slightly more difficult to explain why the participants to a conversation have any responsibility to say something when they do not endorse certain evaluations. On the other hand, the "objective" version of the presuppositional account that I favor has a fairly direct answer: if the presupposed content is about targets, rather than the speaker's state of mind, then everyone is somehow responsible for the evaluation that is taken for granted. Note however that it is not impossible for the "subjective" option to explain complicity. I can think of two lines of explanation, none of which I find completely convincing. The first one is to claim that people are in certain ways responsible for the beliefs endorsed by the people they deal with; in this sense, avoiding challenging the use of a slur would be endorsing the attitude associated to the slur, even if all the slur was doing was to inform about the mental state and dispositions of the slur-user. The second strategy would be to rely on the notion of "presupposition of commonality", that is, the expectation that the conversation participants are all alike under relevant respects. The notion is used by Lopez de Sa (2008: 304) to talk about predicates of taste: "the relevant expression (...) triggers a presupposition of commonality to the effect that the participants of the conversation are all alike in the relevant respects". If we admit an extension from the taste domain to the moral dimension, we would have the following account: a slur like 'wop'

activates the presupposition that "the speaker despises Italians *qua* Italians" and in addition it also triggers the presupposition that the conversation participants feel in the same way about Italians.

3. Conditions sine quibus non

In order for the approach I just sketched to be viable, that is, in order for slurs and thick terms to be analyzed along similar lines, a few issues should be settled. I will consider three aspects that in my opinion have prevented slurs and thick terms from being treated in the same way: (i) the descriptive content of slurs and the notion of "group"; (ii) the wrongness of slurs and (iii) the negativeness of slurs. Finally, I will consider what I take to be the relevant differences between slurs and thick terms.

(i) The target group

First, I would like to dwell on the notion of "group" that seems to be involved when dealing with slurs. We have seen already that slurs are taken to denigrate groups or individuals on the basis of their belonging to a certain class. However, scholars do not usually specify what counts as a target class; instead, they present some examples involving ethnic or geographical origin, religion, gender, sexual orientation. Nunberg (forthcoming) calls these categories "the deep fatalities that have historically been the focus of discrimination or social antagonisms that we see as rents in the fabric of civil society". As a result, there is no consensus whether certain terms, especially those that are not about the listed typical categories, can count as slurs. Take for example "foureyes" for people wearing glasses, or "stinkpotter", for owners of a motorized boat[3]. Nunberg claims that it was a sign of methodological incuriosity on the side of scholars to focus on a restricted set of examples exclusively including the worst slurs like "faggot" and "nigger". Such a restriction put at the center of the debate certain features (the unforgivable wrongness and moral sordidness of epithets) that do not seem to *necessarily* characterize slurs as a class.

> Someone says "That building is full of flacks [publicists]" or "*Mes collègues sont tous fachos*[4]". (...) I don't think many people would want to argue that those utterances aren't truth-evaluable, or that they're purely expressive, or that they're useless[5] to us, all claims that people have made about words like nigger. Not that the claims about the use of that word are wrong—that's another question—but they seem to apply only to words that convey unfounded or indefensible contempt for the members of a racial or ethnic group, which make for a poor candidate for a universal linguistic type (Nunberg forthcoming).

I agree with Nunberg on this and my proposal predicts that pejorative terms like "flack" for "publicist" and "facho" for "fascist", just like "foureyes" and "stinkpotter", can in fact count as slurs, even if the categories they target are not prototypical. This brings up a very important issue concerning the descriptive content of slurs, namely what categories can count as a target group. Scholars have attempted to define what could count as a target group. By observing existing slurs in English, we can already discard the hypothesis that for a property to determine a target group, there must be some kind of self-identification among the people instantiating such a property. It is probably true for properties like "being Italian" or "being

3 "Sailing enthusiasts deprecate the owners of motor craft as 'stinkpotters' but we probably wouldn't call the word a slur" – Nunberg (forthcoming).

4 In Italics in the original.

5 The allusion is to Hornsby (2001), who talks about slurs as useless words.

Jewish" targeted by the slurs "wop" and "kike", but epithets like "gook" do not involve a property with which people can self-identify: "gook" was used to call all natives of the regions occupied by the US army (Nicaraguans, Costa Ricans, Filipinos, Vietnam), who do not identify with each other (see Jeshion, 2016: 135). Hom and May (forthcoming) have formulated three options to establish what counts as a group G, the target of a pejorative:

> I) There are no restrictions on G; it can be instantiated by any group whatsoever. This is in effect the claim that group membership is not something that is morally evaluable.
> II) There is a restriction on G supplied by a theory of natural groups. This theory would isolate racial, religious, gender, sexual orientation, etc. as natural groups, and hence as targets of pejoration.
> III) There is a restriction on G provided by ideologies that are active in sociocultural contexts. A group could be a value of G only insofar as there is a discriminatory cultural norm that supports it (Hom and May forthcoming).

The authors embrace option (III), while keeping the idea expressed in option (I) that group membership is *not* a morally evaluable feature:

> What are the criteria for choices of G such that there will be a pejorative term with the meaning (sense) PEJ(G)? – (...) it is reserved for groups that for whatever odious reasons have associated with them an unjust, hateful or discriminatory ideology that is culturally ingrained within society (*Ibidem*).

I agree with the authors about rejecting option (II) as a non-starter[6], but I do not embrace their solution. In my account *any* property can individuate a target group, as long as there are some speakers who find it interesting or relevant to pick out such a property and convey an evaluation about the objects instantiating it. Once we admit that the target class of a slur could be any (it does not necessarily have to do with nationality, ethnicity etc.), it is easier to conceive a unified analysis of slurs and thick terms. Hybrid evaluatives can trigger evaluations on *any* descriptive property, that is instantiated by people or things.

(ii) The wrongness of slurs

Note that in the above discussion about what counts as a target group there is no requirement on whether the evaluation is appropriate or not. My criterion to define a slur differs from most accounts of slurs in that slurs do not have to be wrong *by definition*, in the sense that it's not part of how I define a slur that the triggered evaluation is unjust. According to Predelli (2010):

> notwithstanding the confused and unpleasant attitude apparently conveyed by uses of slurs of that sort (...) racist and xenophobic attitudes are empirically incorrect: there is no conceptual (and, more importantly, no meaning-encoded) difficulty in supposing that membership in an ethnic or national group provides satisfactory motivation for a hostile attitude (Predelli 2010: 180).

And he adds:

6 About the controversial analysis of notions such as gender and race in terms of natural kinds, see *i.a.* Haslanger (2000), Diaz Leon (2012), Spencer (2012).

> It is a lamentable historical accident that bigots and racists have taken the lead in the production of 'lexically encoded' slurs. Still, exceptions abound: 'pigs' may well have been an eminently usable derogatory term for typical police officers in the sixties (or at least for individuals belonging to certain repressive institutions). Moreover, the statistical prominence of racial, sexist, and xenophobic simple English slurs is easily compensated by the 'expressive compositionality' partially discussed above: 'damn fascist' and 'fucking racist' are as subjectively expressive and linguistically non-defective as xenophobic slurs, but presumably eminently usable by at least some readers of this essay (*Ibidem*: 184).

I leave room for the possibility that certain hybrid evaluatives might trigger an appropriate evaluation; however, this mainly concerns certain non-objectionable thick terms such as 'courageous', presumably not slurs[7]. So I shall in a sense endorse Hom and May's claim that the properties that slurs pick out are *typically* not morally evaluable (being Italian, being homosexual, etc.), but I do not maintain that it holds for every hybrid evaluative: I am with Predelli in saying that the fact that certain hybrid evaluatives (most slurs) target people on the basis of incorrect assumptions is a "lamentable historical accident", not an essential feature of evaluatives in general. In other words, it is not *essential* to slurs that they presuppose unjust and unjustifiable evaluations, only that they convey evaluations. In the case of thick terms, it is particularly clear that the appropriateness is a variable parameter. As a matter of fact, scholars distinguish objectionable and non-objectionable thick terms (Eklund 2011, 2013, Harcourt and Thomas 2013, Kyle 2013, Väyrynen 2013): in the former case, speakers endorse the conveyed evaluation; in the latter, they do not.

(iii) The negativeness of slurs

Lastly, slurs do not have to convey a necessarily negative "hateful" evaluation; it's conceivable that a slur might convey an evaluation with a positive polarity. An example of what a slur with a positive polarity might look like is "Aryan": it was used by Nazis to talk about Indo-Europeans, while conveying some positive evaluation, supporting the idea that being Indo-European is good in itself. While it is a debated topic for slurs, it is quite clear that thick terms can convey a positive evaluation: it is the case for "chaste", "courageous", etc. Note that the question of polarity is completely orthogonal to the issue of objectability: for example, a speaker may well share the negative evaluation conveyed by "brutal" while rejecting the negative evaluation conveyed by "lewd" and, on the other hand, one might endorse the positive evaluation conveyed by "generous", but reject the one conveyed by "chaste". Unlike the objectionable/non-objectionable distinction, the polarity of the evaluation is a lexically encoded feature, independent of the set of values endorsed by speakers.

In sum, once we acknowledge that the target class of a slur could be any (it does not necessarily have to do with nationality, ethnicity etc.), and in principle also the polarity of the evaluation could be any and it could be appropriate or not, it is easier to conceive a unified analysis of slurs and thick terms. Yet, what needs to be stressed is that slurs are usually both unjust and negative, as they typically convey a negative evaluation of their targets on the basis of characteristics that in no way ground or justify derogation, exclusion etc. However, I shall underline once more that the basic linguistic mechanism through which slurs encode evaluative content does not *require*

7 An anonymous reviewer of Phenomenology and Minds has suggested to me a very interesting comparison to the possibility for non-verbal pejoration to be morally justified, such as the depiction of the Nazis as pigs in the graphic novel *Maus* by Art Spiegelman.

that the evaluation conveyed by the slur is unjust, nor that it is negative.
I take the crucial differences between slurs and thick terms to concern their descriptive content. In the case of slurs, the descriptive content amounts to predicates like "Italian", "German", "homosexual", etc., that are not typically context-sensitive. In the case of thick terms, on the contrary, we find gradable adjectives (see *i.a.* Kennedy and McNally 2005; Kennedy 2007). Take for example "lewd", that in my analysis would mean something like "sexually explicit" while triggering an evaluative presupposition. What counts as "sexually explicit" varies from context to context. It is reasonable, for example, to suppose that what counts as "sexually explicit" in a monastery is different from what counts as "sexually explicit" at a Carnival party: different thresholds on possibly different scales. Such gradability of the descriptive content gets inherited by the thick term: a "lewd"-user, namely someone who uses the term "lewd", will apply it or not according to contextually determined parameters. On the contrary, slurs tend not to display the same gradability and to typically have a simpler and more determined descriptive content.

4. Reference and extension

One of the most debated question for slurs and thick terms concerns reference: to what do they refer, if they refer *at all*? Consider slurs. Take the following utterance:
(14) Bianca is a wop.
Authors like Christopher Hom and Robert May (Hom and May 2013; 2014; forthcoming) take slurs to be terms with a necessarily empty extension, as they roughly mean something along the lines of "worthy of negative moral evaluation for being G", where "G" refers to a discriminated group (we saw already that for Hom and May belonging to a group cannot under any circumstance ground the negative evaluation conveyed by slurs). Hom and May (forthcoming) suggest an analogy between slurs and fictional terms. So for example "Unicorns are white" is fictionally true, but materially false, whereas "Unicorns do not exist" is fictionally false, but materially true. According to them slurs work in a similar way. Suppose that the person to which "Bianca" refers is Italian. Compare:

(14) Bianca is a wop.
(15) Bianca is a boche.

In the fiction of racism targeting Italians and Germans, (14) is fictionally true, but materially false, whereas (15) is both fictionally and materially false. A problematic prediction of this approach is that "Bianca is not a wop" turns out to be necessarily true, whether the subject is Italian or not.
Other scholars claimed that slurs refer to the same objects as their neutral counterparts. In this case, (14) is true and (15) is false. This view has the controversial consequence that (14) turns out to be true, as well as utterances like "Italians are wops". The defenders of this approach advocate for a distinction between a *technical* sense of "true", that only involves truth-conditions, and a *folk* sense of "true" that takes into consideration pragmatic factors of various kind. In this spirit, "true" should not be confused with "assertable".
The presuppositional analysis offers a fairly direct answer to questions about reference: hybrid evaluatives refer to the same objects as their neutral counterpart, but, being presuppositional triggers, they are felicitous only if the evaluation they convey is correct. The presuppositional analysis accounts for the following intuition: if one does not endorse a certain moral perspective involved by a hybrid evaluative, an utterance featuring the term is not felicitous nor clearly evaluable. Consider:

(16) Is Bianca a wop?
(17) Is *La vie d'Adele* lewd?

A speaker that does not endorse the evaluations triggered by "wop" and "lewd", would not really know how to answer, as both a yes-answer and a no-answer would equally take for granted the presupposed content. In order to provide an answer at the descriptive level without accepting the presupposition, they might say something like "She is Italian, but there is nothing wrong with that" to (16) and something like "It is sexually explicit, but there is nothing bad in that" to (17).
Note that it is again crucial what is exactly presupposed here. Recall the two options: the "objective" option about the target class ("Italians are despicable because of being Italian") and the "subjective" option about the speaker ("the agent believes that Italians are despicable because of being Italian"). I favor the latter, which provides a better account for complicity. However, while the "subjective" presupposition is true whenever the speaker truly endorses the relevant attitudes, the "objective" option brings up a very complex question: what does it mean for an evaluative presupposition to be correct? Presumably, that the evaluation it conveys is correct. But how do we establish whether "Italians are bad for being Italians" is correct in the context of utterance? The answer to these questions crucially depends on what theory of value one favors. One of the most promising approaches comes from the fitting attitude theories (FA), that define "good" as "fitting object of a pro attitude" (Ewing 1947: 152), where "pro attitude" covers "any favorable attitude to something" (*Ibidem*: 149), such as "choice, desire, liking, pursuit, approval, admiration". Similarly, "(...) what is evil [is] a suitable object of anti-attitudes" (Ewing 1939: 9). However, the question of *how* to establish whether certain sentiments of approbation (or reprobation) are warranted remains unanswered and it is probably beyond the scope of the present work. For the time being, I will leave this "hot potato" to philosophers working in ethics.
Note, though, that these issues do not only concern ethicists, as they have important consequences for the theory of presuppositions in general. We say that for an utterance to be felicitous, its presuppositions must be true. However, since the presupposition can be evaluative rather than descriptive, either we add a clause that says "*when the presupposition is descriptive*", or we change the truth requirement into something like "appropriateness" or "correctness". Again, we still face the challenge of determining when a certain evaluation is correct[8]. In a nutshell: the interesting difference is between the objectionable and nonobjectionable hybrid evaluatives, i.e. between the evaluatives that convey appropriate evaluations and those that do not (namely, slurs and objectionable thick terms), rather than between slurs and thick terms. The criterion to establish whether a presupposition is correct (thus, whether an utterance featuring a hybrid evaluative is felicitous or not), depends on what kind of moral theory one endorses, which is still an open question. However, I hope I clarified a bit the theoretical options at stake.

5. The big picture

Part of the interest for evaluatives on the side of scholars in philosophy of language, linguistics and metaethics is due to the peculiar hybrid nature of these terms and to properties such as projection and the like. However, a bigger picture emerges from the linguistic analysis of these terms that I presented. Something about the *function* and the potential of hybrid evaluatives. As we have seen, my analysis of hybrid evaluatives relies on presuppositions, which are – in Chilton's words – "at least one micro-mechanism in language use which contributes to the

8 I thank Uriah Kriegel for helpful and insightful discussion about this topic.

building of a consensual reality". By employing these terms, we implicitly take for granted a certain moral perspective, certain sets of beliefs concerning what is good and what is bad. We implicitly apply a certain lens to the world and expect everyone else to do the same. As we noted in discussing complicity, because the presupposed content is presented as not open to discussion, if it is not objected it has the potential to shape contexts. In this sense, using hybrid evaluatives is a powerful tool through which language not only encodes evaluation, but is also able to impose it. Talking about the stereotypes evoked by slurs, Nunberg appeals to the notion of "shortcut": "Stereotypes, negative and positive, are among the cognitive shortcuts we rely on to make sense of the world and to guide our responses to it." I argue that this is true not just for stereotypes, but for hybrid evaluatives in general, as they are devices through which language can convey evaluations in a way that is both linguistically encoded and implicit.

REFERENCES

Camp, E. (2013), "Slurring Perspectives", *Analytic Philosophy*, Vol. 54, No. 3, pp. 330-349;
Cepollaro, B. (2015), "In Defense Of A Presuppositional Account Of Slurs", *Language Sciences*, Vol. 52, pp. 36-45;
Cepollaro, B. and Stojanovic, I. (2016), "Hybrid Evaluatives", *Grazer Philosophische Studien*, Vol. 93, No. 3, pp. 458-488;
Chierchia, G. and McConnell-Ginet, S. (1990), *Meaning and Grammar: An Introduction to Semantics*, The MIT Press, Cambridge MA;
Chilton, P. (2004), *Analysing Political Discourse: Theory and Practice*, Routledge, London and New York;
Croom, A. (2011), "Slurs", *Language Sciences*, Vol. 33, pp. 343–358;
Diaz-Leon, E. (2012), "Social Kinds, Conceptual Analysis, and the Operative Concept: A Reply to Haslanger", *Humana.Mente Journal of Philosophical Studies*, Vol. 22, pp. 57-74;
DiFranco, R. (2014), "Pejorative Language", *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/pejorati/;
Eklund, M. (2011), "What are thick terms", *Canadian Journal of Philosophy*, Vol. 41, No. 1, pp. 25-49;
Eklund, M. (2013), "Evaluative Language and Evaluative Reality", In Kirchin, S. (ed.), *Thick Concepts*, Oxford University Press, Oxford;
Ewing, A. C. (1939), "A Suggested Non-Naturalistic Analysis of Good", *Mind*, Vol. 48, pp. 1-22;
Ewing, A. C. (1947), *The Definition of Good*, MacMillan, London;
Gibbard, A. (1992), "Thick Concepts and Warrant For Feelings", *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 61, pp. 267-283;
Harcourt, E. and Thomas, A. (2013), "Thick Concepts, Analysis, and Reductionism", in S. Kirchin (ed.), *Thick Concepts*, Oxford University Press, Oxford, pp. 20-43;
Haslanger, S. (2000), "Gender and race:(What) are they?(What) do we want them to be?", *Noûs*, Vol 34, No. 1, pp. 31-55;
Hom, C. and May, R. (2013), "Moral and Semantic Innocence", *Analytic Philosophy*, Vol. 54, No. 3, pp. 293-313;
Hom, C. and May, R. (2014), "The inconsistency of the identity thesis", *Protosociology*, Vol. 31, pp. 113-120;
Hom, C. and May, R. (forthcoming), "Pejoratives as Fiction", in D. Sosa (ed.) (forthcoming), *Bad words*, Oxford University Press, Oxford;
Hornsby, J. (2001), "Meaning and Uselessness: How to think about derogatory words", *Midwest Studies in Philosophy: Figurative Language*, Vol. 25, pp. 128-141;
Jeshion, R. (2013), "Expressivism and the Offensiveness of Slurs", *Philosophical Perspectives*, Vol. 27, No. 1, pp. 231-259;

Jeshion, R. (2016), "Slur Creation, Bigotry Formation: the Power of Expressivism", *Phenomenology and Mind*, 11, pp. 130-139;
Kennedy, C. (2007), "Vagueness and grammar: The semantics of relative and absolute gradable adjectives" *Linguistics & Philosophy*, Vol. 30, pp. 1-45;
Kennedy, C. and McNally, L. (2005), "Scale structure, degree modification and the semantics of gradable predicates", *Language*, Vol 81, pp. 345-381;
Kyle, B. G. (2013), "How Are Thick Terms Evaluative?", *Philosophers' Imprint*, Vol. 13, No. 1, pp. 1-20;
López de Sa, D. (2008), "Presuppositions of commonality: An indexical relativist account of disagreement", in M. García-Carpinterol, and M. Kölbel (2008), *Relative truth*, Oxford University Press, Oxford, pp. 297-310;
Nunberg, G. (forthcoming), "The Social Life of Slurs", in D. Fogal, D. Harris & M. Moss, (eds.), *New Work on Speech Act*, Oxford University Press, Oxford;
Predelli, S. (2010), "From the Expressive to the Derogatory: On the Semantic Role for Non-Truth-Conditional Meaning", in S. Sawyer (ed.), *New Waves in Philosophy of Language*, Palgrave-MacMillan, New York, pp. 164-185;
Spencer, Q. (2012), "What 'biological racial realism' should mean", *Philosophical studies*, Vol. 159, No. 2, pp. 181-204;
Richard, M. (2008), *When Truth Gives Out*, Oxford University Press, Oxford;
Schlenker, P. (2007), "Expressive presuppositions", *Theoretical Linguistics*, Vol. 33, No. 2, pp. 237-245;
Simons, M., Tonhauser, J., Beaver, D. and Roberts, C. (2010), "What projects and why", in N. Li, and D. Lutz (eds.), *Semantics and linguistic theory*, (SALT) 20, CLC Publications, Ithaca, NY, pp. 309-327;
Väyrynen, P. (2009), "Objectionable Thick Concepts in Denials", *Philosophical Perspectives*, Vol. 23, pp. 439-469;
Väyrynen, P. (2013), *The lewd, the rude and the nasty: A study of thick concepts in ethics*, Oxford University Press, Oxford.

FRANCESCA PANZERI
University of Milano-Bicocca
francesca.panzeri@unimib.it

SIMONE CARRUS
Vita-Salute San Raffaele University
carrussimone@gmail.com

SLURS AND NEGATION

abstract

We present the results of an experimental study that aims at establishing whether the offensive component of slurs exhibits nondisplaceability (Potts 2007). We found that the derogatory content survives in conditionals and questions (supporting a pragmatic approach), and diminishes in indirect reports (in line with presuppositional accounts); surprisingly, the offensiveness of slurs results almost nullified in negated sentences. In a second study, we explore the hypothesis that negated slurs were rated as not offensive because the negation was interpreted as metalinguistic.

keywords

slurs, pejoratives, metalinguistic negation, semantics/pragmatics interface

Phenomenology and Mind, n. 11 - 2016, pp. 170-180
DOI: 10.13128/Phe_Mi-20117
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Slurs are derogatory epithets that target specific groups, identified mainly on the basis of race (*nigger* for a black person), nationality (*wop* for Italian), religion (*kike* for Jew), sexual orientation (*faggot* for gay). Slurs differ from other pejoratives (*moron*, *asshole*) because they insult a person inasmuch as (s)he belongs to a specific group, that can be identified by means of a not-offensive expression, the neutral counterpart (or non-pejorative correlate).
Slurs are particularly hateful and pernicious because they convey and reinforce stereotypes about the target group, they harm "their target's self-conception and self-worth, often in ways that are common to the social group as a whole" (Jeshion 2013: 314), and they are thus considered taboo, prohibited words (Anderson & Lepore 2013b).
A lively debate has developed in recent years regarding the *meaning* of these expressions, and, in particular, the proper way to account for their offensiveness. A first distinction can be drawn between deflationary accounts and content-based theories: the former view slurs' offensiveness as stemming not from a derogatory *content* they would express, but from other extra-linguistic factors, such as the presence of relevant edicts surrounding their prohibition (Anderson & Lepore 2013b), or the fact slurs are tools to signal the affiliation with racist views (Numberg 2013). Content-based approaches, on the other hand, aim at offering an analysis of slurs offensive *content*, in semantic or pragmatic terms. Semantic theories assume that a slur encodes as part of its literal meaning the derogation with respect to the target group. A sentence such as (1), then, is considered to be truth-conditionally equivalent to (something like) (2):

(1) Leo is a faggot.
(2) Leo is homosexual and despicable because of it.

Pragmatic accounts view the reference to the target group and the derogatory attitude as two separate components of a slur: the sentence in (1) is seen as semantically equivalent to (3), that constitutes the descriptive content, whereas the derogation with respect to the target group, paraphrasable as in (4), is pragmatically conveyed, as a presupposition (Schlenker 2007, Cepollaro 2015), or conventional implicature (Potts 2007, Williamson 2009, McCready, 2010):

(3) Leo is a (male) homosexual.
(4) Homosexuals are despicable.

In this contribution, we focus on the predictions these content-based theories[1] make about the survival of slurs' offensive component in various linguistic contexts. One of the slurs' main peculiarities, in fact, is that they appear to be offensive not only when they are used in a positive statement, as in (1), but also when they are negated, as in (5):

(5) Leo is not a faggot.

Notice that this feature distinguishes slurs from simple pejoratives: both (5) and (6) can be viewed as rude, nevertheless it is only the former that sounds derogatory towards the target group of homosexuals, whereas (6) should not carry any offense.

(6) Leo is not an asshole.

Semantic theories assume that the derogatory content of a slur is part of its literal meaning, and therefore it falls under the scope of operators such as negation, conditionals, questions. In other words, in a sentence like (5), the negation operates on (and thus denies) the derogation – which would not be asserted. Hom admits that "[f]or many, the taboo surrounding epithets is not limited to their direct use, but covers their occurrence within quotation, fiction, intensional contexts, questions, negations, conditional antecedents, and even extends to phonologically similar, but semantically distinct, expressions" (Hom 2008: 427), but he proposes to distinguish between *real* derogation, stemming only from the actual predication (that is, only when the slur is attributed to someone, as in (1)), and *squeamishness*, a feeling of discomfort when a slur is embedded in a non-predicative environment (as in (5)). According to Hom, in other words, the actual derogation towards the individual Leo is indeed present only in (1), whereas a sentence like (5) does not derogate Leo, even if the fact that the speaker chose to use a slur (instead of the neutral counterpart) might suggest that she shares bigots stereotypes and contempt towards homosexuals: "uses of epithets often carry the presumption that its speaker subscribes to the underlying racist institution" (Hom 2008: 435).
Pragmatic theories, on the other hand, expect only the truth-conditional content (3) to fall within the domain of linguistic operators, whereas the derogatory component (4), being pragmatically conveyed, could survive and be transmitted to the whole sentence. Thus, within a pragmatic approach to slurs, the derogation towards the target group expressed by means of a slur is predicted to be the same in positive contexts such as (1) and negative statements like (5).
And, in fact, it has been claimed that slurs exhibit non-displaceability (Potts 2007), that is, their derogatory content "scopes out" (Hedger 2012) of different linguistic contexts. Williamson (2009: 146), for instance, considering the occurrences of *boche* (slur for Germans) notes that "the xenophobic abuse is preserved in the negations". Enlarging the perspective, Croom (2014: 228) acknowledges that "the potential offensiveness of slurs is [...] further evidenced by more straightforwardly linguistic considerations, such as through an analysis of their projection behavior across a diverse range of linguistic contexts". And following the same intuition, Hedger (2012: 74) proposes the following generalization: "the offensive content of slurs scopes out of logical operators".
In other words, slurs' offensiveness would be conveyed not only in predicative contexts (like (1)) and negative statements (like (5)), but also in sentences where the slur is hypothesized (as in (7)), questioned (as in (8)):

1 We will not take into consideration deflationary accounts, since, as already alluded to, they do not view slurs as expressing a derogatory *content*, and thus it is hard to pin down their predictions in specific contexts.

(7) If Leo is a faggot, he knows the answer.
(8) Is Leo a faggot?

Although we can find in the literature several different ways to refer to this projection behaviour ("nondisplaceability", Potts (2007); "scoping-out", Hedger (2012); "project-out", Camp (2013); "persistence of offensiveness", Hom and May (2013)), the phenomenon has been defined with sufficient clearness. Indeed, it is supposed to be so revealing for the slurring phenomenology that some researchers (e.g. Jeshion 2013, Cepollaro 2015) use it as a tool to dismiss some semantic accounts (e.g. Hom 2008), in favour of pragmatic approaches.
There is a controversy as to what happens when a slur is embedded in a report, as in (9):

(9) Gianni said that Leo is a faggot.

According to some scholars (Potts 2007), the person (the speaker of (9)) who is reporting what was said by someone else (Gianni) is offensive, since she did not choose to utter a neutral term. Other scholars (Kratzer 1999, Schlenker 2007) maintain that reporting a "bad word" uttered but someone else does not necessarily require the speaker to share the same negative attitude. Schlenker (2007)'s example in (10) illustrates this situation:

(10) I am not prejudiced against Caucasians. But John, who is, thinks that you're the worst honky he knows.

As we already mentioned, pragmatic approaches defend the idea that the derogatory component is not part of what is literally said, rather it is transmitted either as a presupposition or as a conventional implicature. Presuppositional and conventional implicature approaches easily account for the survival of the offensiveness of slurs in the linguistic contexts of negation, antecedents of conditionals and questions, but, quite interestingly, they make different predictions for indirect reports such as (9). Since the indirect report in (9) is a "plug" (Karttunen, 1973), presuppositional approaches to slurs predict that the derogatory import would be blocked in (9), where the offensiveness carried out by the slur *faggot* should be attributed to Gianni, but not (necessarily) to the speaker who reports his words. Theories that view slurs as conveying offensiveness as a conventional implicature, on the other hand, would predict that also the person who reports a slur (uttered by another person) is being offensive towards homosexuals, since conventional implicatures exhibit independence (Potts 2005).
We believe that the intuitions about slurs' scoping out behaviour can be checked by means of experimental studies. There have been several studies in the area of Social Psychology, whose aim was to verify the (perlocutionary) effects of utterance of slurs on the targeted groups (see, a.o., Carnaghi & Maas 2007 and Fasoli *et al.* 2013). Nevertheless, to our knowledge, the *semantic* questions and intuitions about slurs' non-displaceability have not been tested yet (Spotorno & Bianchi 2015).

2. Study 1: Slurs' offensiveness in linguistic contexts

We carried out a study whose principal goal is to establish the offensiveness of slurs in the linguistic contexts of negation, antecedent of conditionals, questions and indirect reports, and compare it to the perceived offensiveness of slurs in isolation. In particular, semantic theories of slurs would predict that the derogative force of slurs is suspended in these linguistic contexts; pragmatic approaches predict that slurs' offensiveness survives under negation, in the antecedent of conditional, and in questions; presuppositional accounts of slurs expect slurs' offensiveness not to be (necessarily) attributed to the person who reports a slur, whereas

accounts that view slurs as conventionally implicated predict the same level of offensiveness also in indirect reports.

2.1. Method

We tested 132 (90 F) Italian undergraduate students, who volunteered to participate to the study, with a mean age of 23;5. The study consisted of four versions of a written questionnaire. The questionnaire comprised two parts. In the first one (Baseline), participants had to rate, on a 7-points scale, the offensiveness of 32 words presented in isolation. The instructions that were given to the participants are reported in (11):

(11) We are interested in the study of the offensiveness of certain expressions. We are going to present you a list of words. For any of these words, please indicate how much you think that that word is offensive, using a scale that goes from 1 ("not at all offensive") to 7 ("extremely offensive").

The 32 words that were presented consisted in 8 slurs (SL); their 8 neutral counterparts (NC); 8 neutral/positive controls (PC); 8 bad words (BW), presented in a pseudorandom order. The items are presented in Table 1, with the Italian items used in the experiment, followed by their English translation or explanation.

SLURS (SL)	NEUTRAL (NC) COUNTERPARTS	POSITIVE CONTROLS (PC)	BAD WORDS (BW)
Frocio (*faggot*)	Omosessuale (*male homosexual*)	Oculista (*eye doctor*)	Coglione (*asshole*)
Negro (*nigger*)	Di colore (*colored*)	Castano (*hair brown*)	Testa di cazzo (*dickhead*)
Ciccione (*fatso*)	Sovrappeso (*overweight*)	Mancino (*left-handed*)	Stronzo (*turd*)
Musogiallo (*Chink*)	Cinese (*Chinese*)	Inquilino (*tenant*)	Bastardo (*bastard*)
Handicappato (*handicapped*)	Diversamente Abile (*disabled*)	Studente (*student*)	Deficiente (*moron*)
Terrone (slur for *Southern Italian*)	Meridionale (*Southern Italian*)	Adulto (*adult*)	Imbecille (*imbecile*)
Zitella (*spinster*)	Nubile (*nubile*)	Geometra (*surveyor*)	Idiota (*idiot*)
Crucco (*Kraut*)	Tedesco (*German*)	Ciclista (*cyclist*)	Stupido (*stupid*)

Table 1. List of the 32 items of the Baseline, divided in the four categories of slurs (SL), their neutral counterparts (NC), positive controls (PC), and bad words (BW)

In the second part of the questionnaire, Linguistic Context, participants were asked to rate, always on a 7-points scale, the offensiveness of a person who utters a sentence that contain a word (SL/NC/PC/BW) embedded under the linguistic contexts of negation (NEG), antecedent of a conditional (ANT), question (QUE), and indirect report (IND). The instructions given to the participants are provided in (12), and one example is given in (13).

(12) Now, we ask you to do something slightly different. Imagine that you are, by chance, overhearing a conversation amongst persons you do not know. We ask you to indicate

how much you think that **the person who is speaking** (that will be specified every time) has been offensive – always using a scale that goes from 1 ("not at all offensive") to 7 ("extremely offensive").

(13) **Claudio**: "Lino is not a faggot"

How much do you think that **Claudio** has been offensive?

1	2	3	4	5	6	7
Not at all offensive						Extremely offensive

2.2. Results Results indicate that, in the Baseline, slurs and bad words are perceived as offensive, and overall the offensiveness of slurs and bad words is not statistically different. Neutral counterparts and positive controls, on the other hand, are not considered to be offensive. The results of the Baseline part of the questionnaire are plotted in Figure 1.

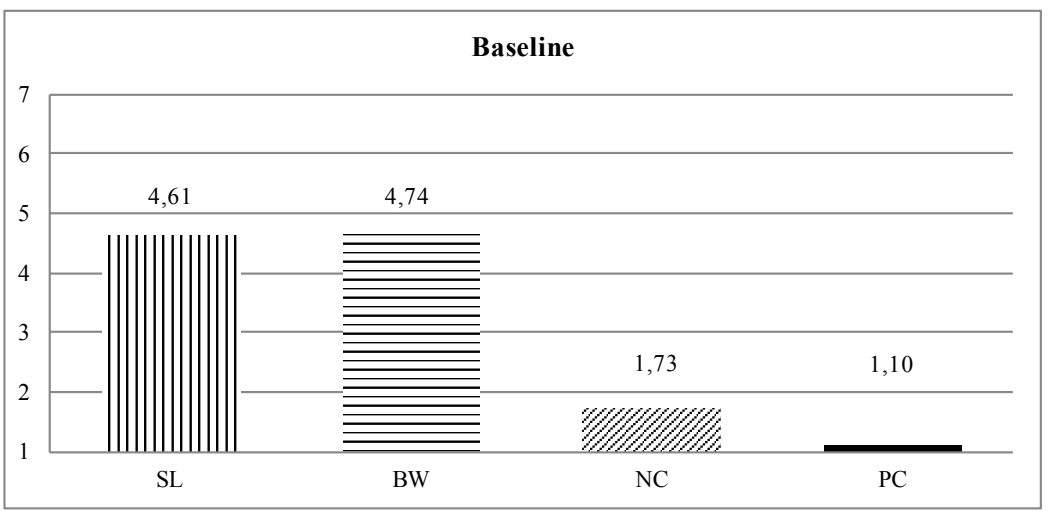


Figure 1. Average of the perceived offensiveness (on a 7-point scale) of the 32 words presented in isolation, grouped in the categories of slurs (SL, vertically striped), their neutral counterparts (NC, diagonally striped), positive controls (PC, black), and bad words (BW, horizontally striped).

We compared the offensiveness of slurs, bad words, neutral counterparts and positive controls in the baseline, to the perceived offensiveness of a person who utters a sentence that contains the same items, embedded in the linguistic contexts of negation (NEG), antecedent of conditionals (ANT), question (QUE) and indirect report (IND). We found that a person uttering a slur is perceived as being offensive even if the slur is embedded in a question or in the antecedent of a conditional; a person who reports the statement of someone else who used a slur is herself perceived as being offensive, but to a lesser degree; quite surprisingly, when a person utters the negation of a statement that contains a slur, she is not perceived as being particularly offensive. The degree of offensiveness of the slurs

in isolation (baseline) and of the person uttering a slur in the aforementioned linguistic contexts are reported in Figure 2.

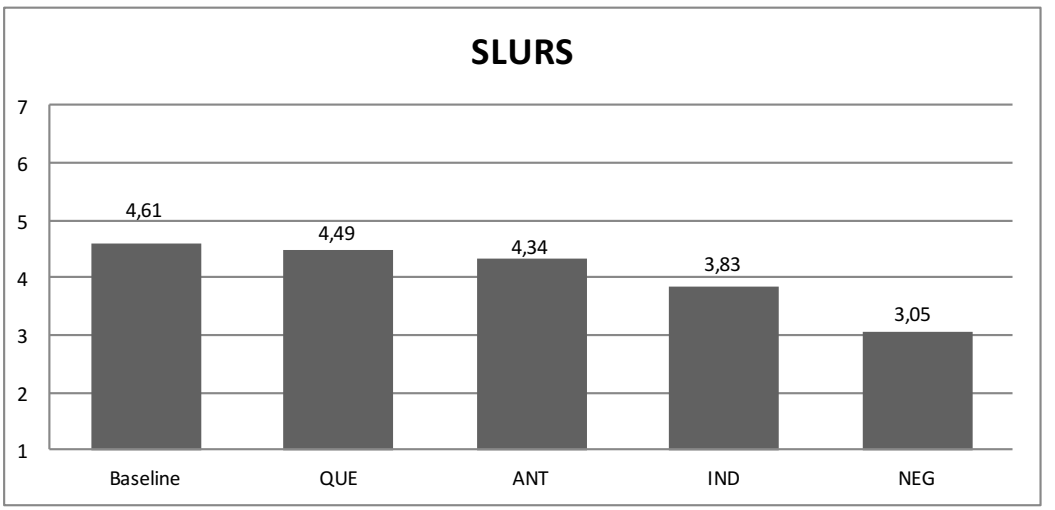


Figure 2. Average of the perceived offensiveness (on a 7-point scale) of the slurs in the baseline and in the linguistic contexts of questions (QUE); antecedent of conditionals (ANT); indirect reports (IND), and negation (NEG).

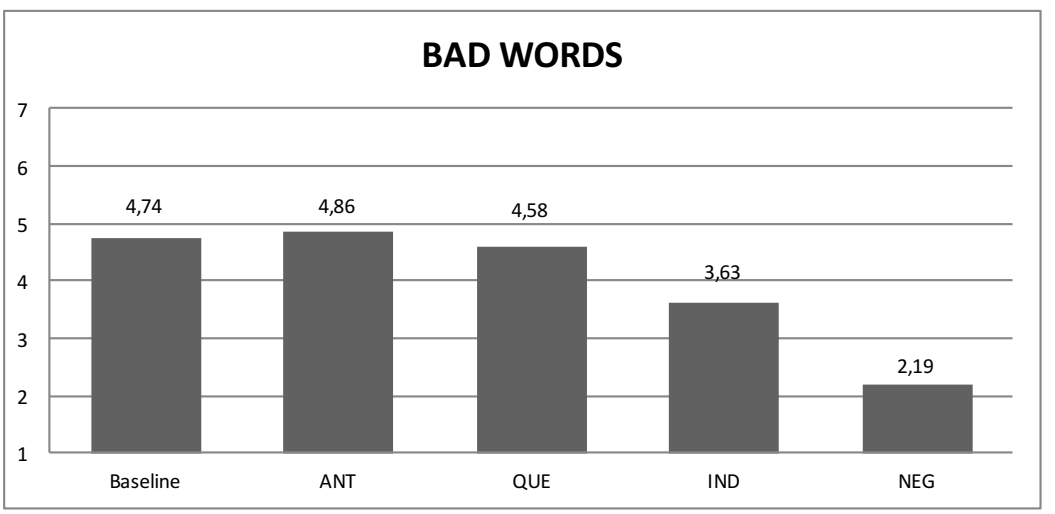


Figure 3. Average of the perceived offensiveness (on a 7-point scale) of the bad words in the baseline and in the linguistic contexts of antecedent of conditionals (ANT); questions (QUE); indirect reports (IND), and negation (NEG).

Bad words exhibit a somehow parallel behaviour, with their offensiveness scoping out from the linguistic contexts of antecedents of conditionals and questions, with indirect reports that

diminish, but not cancel the derogation of these words, and, again, with negation that almost nullifies the offensiveness of bad words. The results are presented in Figure 3.

2.3. Discussion

We found that, as stated in the literature, the derogatory import of slurs scopes out from the linguistic contexts of antecedents of conditionals and questions, since a person who utters a slur that is embedded in one of these contexts is perceived as offensive as the slur in isolation. We did not get clear results on what happens in indirect reports, since the degree of offensiveness of a person who reports a slur uttered by someone else is perceived offensive to a lesser degree. Within a presuppositional approach to the meaning of slurs, this might be interpreted as indirect evidence that *verba dicendi* do constitute plugs (that is, they block the offensiveness of the slur, that is thus attributed only to the person who actually uttered the slur), and that the person who reports the slur uttered by someone else is felt to be less accomplice of the choice of that word.
The result we obtained for the linguistic context of negation is quite surprising, since all theories of slurs agree that negating a slur does not have an effect on its offensiveness, or, at least, that the level of perceived derogation (or squeamishness, in Hom's terms) should be the same as the one attested for conditionals and questions. In our experiment, on the other hand, we found that participants were rating the offensiveness of a person who utters a negation of a slur, as in the example given in (13), as only slightly offensive. Before discussing the results in the light of the theoretical approaches to slurs, we first need to provide a possible explanation for the unexpected result for negation.
We hypothesize that our participants might have interpreted the speaker's statement as a metalinguistic negation: asserting a negated proposition is an appropriated move when the purpose is to contradict what is (salient or) asserted by someone else; in that case, the negated statement could be interpreted as a correction of the expressive, derogatory component of the corresponding affirmative statement ("Marco is not a faggot, he is a homosexual"). We decided to test this hypothesis with a new test in which we elicit possible continuations of negated statements containing slurs, neutral counterparts, bad words, and positive controls.

3. Study 2: Negating a slur

In this second study, we presented our participants with a negated statement, and we asked them to provide a continuation of that sentence. Our purpose was to test whether, and to what extent, the negation of a slur could be interpreted as a metalinguistic negation.

3.1. Method

We tested a total of 100 Italian undergraduate students (76 F), with a mean age of 19;9 (18-52). The study consisted of two different versions of a written questionnaire. Participants had to provide a continuation of a total of 16 negated sentences. The instructions that were given to the participants are reported in (14):

(14) We ask you to complete the following sentences, in such a way that they make sense. For instance, the following sentence:
A. It is not raining in Milan, ...
could be completed with these continuations:
A1. It is not raining in Milan, it is sunny.
A2. It is not raining in Milan, it is pouring down.
A3. It is not raining in Milan, but in Rome it is.

Please note that one of the three possible continuations provided as an example, the one in A2, does constitute a metalinguistic negation (not raining, but pouring down), where we object to the choice of a less informative term (rain) and correct it with a stronger one (pour).

Nevertheless, this metalinguistic negation has to do with the strength of informativeness required, and not with the offensiveness of the terms.
The test items were the negation of 8 slurs, the same as those tested in the first study, except for the slurs *handicappato* ("handicapped") and *ciccione* ("fatso"), that have been substituted with two other slurs: *sbirro* ("screw", "pig"), the Italian offensive term to refer to police agents, and *strizzacervelli* ("headshrinker"), the pejorative version of psychoanalyst. We added as fillers the negation of 24 more items (the 8 neutral counterparts of the slurs, plus the positive controls and bad words used in the first study). We thus obtained a total of 32 negated sentences, which were divided in two different lists, containing 16 sentences each, in order to ensure that each list contained a slur but not its neutral counterpart.

3.2. Results and discussion

The results were coded by two independent researcher, with the following criterion: the continuations were coded CON whenever they were negating the *content* of the predicate corresponding to the neutral counterpart; as NON-CON when the negation was referring to other aspects, but not involving the content; as IRR when the continuations were irrelevant. Examples of the first case CON are given in (15), provided as possible continuations of the sentence "Leo is not a faggot":

(15) a. He has a girlfriend.
b. He is just a little bit effeminate.

Please note that in both these cases the participant interpreted the negation in "Leo is not a faggot" as the denial of the fact that Leo belongs to the set of homosexuals.
Examples of the second class of NON-CON continuations are given in (16), always following the negated sentence "Leo is not a faggot":

(16) a. He is homosexual.
b. I don't like these expressions.

In those cases, the participants were not interpreting the negation as denying the fact that Leo belongs to the set of homosexuals, but they were objecting to the choice of the slurring expressions, and explicitly correcting that choice.
When the sentence to be continued contains a slur, the overall percentage of NON-CON continuations (that do not deny that the subject belong to the target group) is 33%, with a rather high variability: negated sentences with the slurs *musogiallo* ("chink"), *frocio* ("faggot"), and *negro* ("nigger") are followed by more than 40% of NON-CON continuations (49%, 47% and 40%, respectively); the percentage of NON-CON continuations is between 20% and 30% for the slurs *strizzacervelli* ("headshrinker", 30%), *sbirro* ("screw", 27%), *zitella* ("spinster", 27%), *crucco* ("Kraut", 23%) and *terrone* (derogatory term for Southern Italian, 22%).
Even if these data are preliminary, there is an extremely high percentage of continuations that interpret the negation as referring not to fact that the subject belongs to the target group (the neutral counterpart), but to the offensiveness of the term (of the slur), especially for those terms that were rated as most offensive in the previous study. We believe that these continuations constitute interpretations of the preceding negated sentence as instances of a metalinguistic negation.
Coming back to the surprising results of the first study, we think that the fact that a person who was uttering a sentence of the form "x is not *slur*" was not considered as being offensive, could be explained assuming that the participants were interpreting that sentence as an instance of a metalinguistic negation, that is not indeed offensive.

4. Conclusions

Assuming that at least some of the participants to the first study rated a person who was uttering the negation of slurs as not offensive because they were interpreting the sentence as an instance of a metalinguistic negation, we can conjecture that even "ordinary" negated slurs are perceived as offensive.

Coming back to the predictions made by the various approaches to the meaning of slurs, we believe that our results favour a pragmatic account on slurs, and, in particular, the fact that a person who reports a slur is perceived as less offensive than a person who uses a slur in a conditional or in a question is more easily accounted for within a presuppositional account on slurs.

For reasons of space, we could not further explore the parallelism between slurs and simple pejoratives: it is quite surprising that pejoratives seem to exhibit a behaviour analogous to slurs when they are embedded within a conditional, a question and an indirect report, whereas when they are negated they lose their offensiveness – a fact for which we do not have a principled explanation for.

We are aware that the studies we presented present some shortcomings: the participants are not at all representative (they are all undergraduate students, mostly from the same area, the northern part of Italy); more importantly, the experimental design presented the sentences to be evaluated as offensive out of context: the motivation behind this choice was to focus the participants' attention only to the perceived offensiveness of the slur *per se*, in the various contexts. Nevertheless, we are aware of the fact that there are in-group uses of slurs, and that slurs can also be used in an ironic or metaphorical sense, and we couldn't control for our participants' actual interpretation of the intended context in which slurs were used. Still, we believe that scholars' intuitions about the offensiveness of slurs need to be checked with experimental studies.

REFERENCES

Anderson, L., and Lepore, E. (2013a), "Slurring Words", *Nous*, 47(1), pp. 25-48;
Anderson, L., and Lepore, E. (2013b), "What did you call me? Slurs as prohibited words", *Analytic Philosophy*, 54(3), pp. 350-363;
Carnaghi, A., and Maass, A. (2007), "In-Group and Out-Group Perspectives in the Use of Derogatory Group Labels Gay Versus Fag", *Journal of Language and Social Psychology*, 26(2), pp. 142-156;
Cepollaro, B. (2015), "In defence of a presuppositional account of slurs", *Language Sciences,* 52, pp. 36–45;
Croom, A. M. (2011), "Slurs", *Language Sciences*, 33(3), pp. 343-358;
Croom, A. M. (2013), "How to do things with slurs: Studies in the way of derogatory words", *Language & Communication*, 33(3), pp. 177-204;
Croom, A. M. (2014), "The semantics of slurs: a refutation of pure expressivism", *Language Sciences*, 41, pp. 227-242;
Fasoli, F., Paladino, M. P., and Carnaghi, A. (2013), "Le conseguenze del linguaggio omofobo su omosessuali ed eterosessuali", *Psicologia sociale*, 8(2), pp. 177-192;
Hedger, J. A. (2012), "The semantics of racial slurs: using Kaplan's framework to provide a theory of the meaning of derogatory epithets", *Linguistic and Philosophical Investigations*, 11, pp. 74-84;
Hom, C. (2008), "The semantics of racial epithets", *The Journal of Philosophy,* 105, pp. 416-440;
Hom, C. (2010), "Pejoratives", *Philosophy Compass*, 5(2), pp. 164-185;
Hom, C., and May, R. (2013), "Moral and semantic innocence", *Analytic Philosophy*, 54(3), pp. 293-313;
Hornsby, J. (2001), "Meaning and uselessness: how to think about derogatory words", *Midwest Studies in Philosophy,* 25, pp. 128-141;

Jeshion, R. (2013), "Slurs and stereotypes", *Analytic Philosophy*, 54(3), pp. 314-329;
Karttunen, L. (1973), "Presuppositions of compound sentences", *Linguistic inquiry*, 4(2), pp. 169-193;
Kennedy, R. L. (2008), *Nigger: The strange career of a troublesome word*, Pantheon, New York;
Kratzer, A. (1999), "Beyond ouch and oops: How descriptive and expressive meaning interact", Available from Semantics Archive;
Maciá, J. (2002), "Presuposición y significado expresivo", *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, 17(3), pp. 499-513;
McCready, E. S. (2010), "Varieties of conventional implicature", *Semantics and Pragmatics*, 3, 8, pp. 1-57;
Potts, C. (2007), "The expressive dimension", *Theoretical linguistics* 33(2), pp. 165-198;
Potts, C. (2012), "Conventional implicature and expressive content", In Maienborn, C., von Heusinger, K., and Portner, P. (eds.), *Semantics: An International Handbook of Natural Language Meaning*, DeGruyter 3, pp. 2516–2536;
Schlenker, P. (2007), "Expressive presuppositions", *Theoretical Linguistics*, 33(2), pp. 237-245;
Spotorno, N. and Bianchi, C. (2015), "A plea for an experimental approach on slurs", *Language Sciences*, 52, pp. 241-250;
Williamson, T. (2009), "Reference, inference and the semantics of pejoratives", In Almog, J., and Leonardi, P. (eds.) *The Philosophy of David Kaplan*, Oxford University Press, Oxford, pp. 137-158.

LAURA CAPONETTO
Vita-Salute San Raffaele University
lauracaponetto@gmail.com

SILENCING SPEECH WITH PORNOGRAPHY

abstract

The aim of this paper is to offer a map of the dynamics through which pornography may silence women's illocutions. Drawing on Searle's speech act theory, I will take illocutionary forces as sets of conditions for success. The different types of silencing, I claim, originate from the hearer's missed recognition of a specific component of the force of the speaker's act. In addition to the varieties already discussed in literature (which I label essential, authority, *and* sincerity silencing*), I shall finally consider another kind of silencing produced by the failure to acknowledge the speaker's words as serious (*seriousness silencing*).*

keywords

silencing, speech acts, pornography, misrecognition, disempowerment

Phenomenology and Mind, n. 11 - 2016, pp. 182-191
DOI: 10.13128/Phe_Mi-20118
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. The Silencing Argument Against Pornography

Catharine MacKinnon (1987, 1993) has contended that pornography *silences* women by violating their freedom of speech[1]. Jennifer Hornsby and Rae Langton (H&L) (Hornsby 1993; Langton 1993; Hornsby & Langton 1998) have defended the plausibility of this claim by drawing on Austin's speech act theory. Pornography, H&L argue, spreads false beliefs and expectations about women (*i.e.*, about what they are, what they desire, how they behave in sexual contexts) that interfere with men's capacity to grasp the illocutionary force of certain acts women attempt to perform, thereby causing them to misfire. Silencing, in this frame, is a form of *uptake failure*.

Besides H&L's proposal, other accounts of silencing have recently been provided. As a result, the notion of 'silencing' has become richer but also more elusive. This paper surveys the major conceptions of silencing involved in the debate on the phenomenon. The aim is to offer a map of the ways in which speech – especially women's speech[2] – can be silenced. The discussion focuses on the act of sexual refusal. However, the kinds of silencing I will consider may hinder the performance of other sorts of acts as well. Moreover, though the most interesting views on silencing have been developed in debating the harms of pornography, all varieties of silencing may be brought about[3] by sources other than pornography too (*e.g.*, racist speech). Unlike others who have dealt with silencing, I shall adopt the Searlian framework, which seems better suited to describe the dynamics through which speakers can be silenced, thanks to its high level of accuracy[4]. According to Searle, every illocutionary force can be identified by a set of

1 MacKinnon's allegation narrows its scope to a specific subset of pornographic material intended for a heterosexual male audience and constituted by depictions of women dehumanized as sexual objects and shown as enjoying pain, humiliation or rape. Hereafter, I will use the term 'pornography' to refer to just this subset.

2 Women are the (silenced) speakers I will look at since my discussion focuses on the (silenced) act of sexual refusal. Nevertheless, the types of silencing I examine here may affect other "disempowered speakers" (Hornsby 1995) as well, as they are likely to occur whenever the distribution of power is strikingly unjust.

3 While Hornsby (1993, 2014) suggests that pornography *causes* silencing, Langton (1993) claims that it *constitutes* silencing. I shall remain neutral on whether or not pornography constitutes (rather than merely causes) silencing, since the differences between the two accounts are not relevant for the purposes of this paper.

4 The silencing claim against pornography (especially its "constitutive" version) is highly controversial. As most difficulties raised by critics depend on the Austinian setting endorsed by its supporters, adopting Searle's alternative paradigm may be a good move for eluding (some of) them. In particular, Searle's framework offers some insights to solve the "Authority Problem" (think to the distinction between *power* and *authority* outlined in Searle & Vanderveken 1985: ch. 9). Since the issue falls outside the scope of this paper, I leave the assessment of this hypothesis for another occasion. (For a discussion of the Authority Problem, see Maitra 2012).

conditions for success. In recent literature, we find three main forms of silencing. All of them, I claim, originate from the hearer's failed recognition of a specific component of the force of the speaker's act.
The plan of the paper is the following. After outlining the key features of Searle's analysis of illocutionary force (Section 2), I examine through the Searlian lens four types of silencing (Section 3). Firstly, I discuss what I label *essential silencing*, which involves the audience's failure to recognize the illocutionary point of the speaker's act. It corresponds to H&L silencing (Section 3.1). Secondly, I take into account a type of silencing occurring when the hearer fails to ratify the speaker's authority over a relevant domain (*authority silencing*). Similar sorts of silencing have been discussed by Mary Kate McGowan (2009) and Marina Sbisà (2009) (Section 3.2). Later on, I consider the so-called *sincerity silencing* (McGowan 2014), which occurs when the speaker's utterance is mistakenly taken as insincere (Section 3.3). In addition to these varieties, I then discuss a form of silencing produced by the failure to acknowledge the speaker's words as serious (*seriousness silencing*) (Section 3.4)[5]. Finally, I argue that all these types of silencing are genuine instances of illocutionary disablement, albeit at first glance it might not seem so (Section 4).

2. Searle on Illocutionary Force

As Austin has pointed out, saying something is *eo ipso* doing something. In speaking, we not only utter meaningful expressions (*i.e.*, perform *locutionary* acts), but we perform *illocutionary* and *perlocutionary* acts too. Illocutionary acts correspond to the actions the speaker performs in uttering certain words, or to put it another way, to the peculiar force of the locution in the context of utterance (*e.g.*, *in* saying "Leave him!" the speaker *urged* the hearer to leave him), whereas perlocutionary acts correspond to the effects brought about on the addressee's thoughts, feelings, and behaviours (*e.g.*, *by* saying "Leave him!" the speaker *persuaded* the hearer to leave him). As is known, illocution is the core of the speech act theory. Every use of language has indeed a performative dimension enshrined in its illocutionary force, which turns the uttering of a sentence in the performing of an action.
In attempting to formalize and deepen the notion of 'illocutionary force', Searle & Vanderveken (1985) have defined it as a set of seven success conditions[6]. Every septuple has the following components: (i) *illocutionary point* (the aim an act has in virtue of being an act of a certain kind)[7], (ii) *degree of strength of the illocutionary point* (*e.g.*, differences such as that between asserting and insisting), (iii) *mode of achievement* (the peculiar way in which the illocutionary point must be pursued), (iv) *propositional content conditions* (restrictions imposed on the utterance's content), (v) *preparatory conditions* (states of affairs that must obtain in the world of utterance), (vi) *sincerity conditions* (the psychological state(s) the speaker must have in order to perform a sincere act), (vii) *degree of strength of the sincerity conditions* (*e.g.*, differences such as that between requesting and beseeching).
On my account, illocutionary silencing (in general) is a speech act failure constituted by the systematic and unjust misrecognition by the audience of a specific component of the force of the speaker's act. The term 'misrecognition', as I use it, means simply 'missed (or failed) recognition' (and not 'misinterpretation' or 'wrong recognition'). Such misrecognition constitutes (some sort of) silencing only if it takes place in a (i) systematic and (ii) unjust manner – that is, only if (i) it is brought about by highly widespread beliefs about a certain

5 In Section 3.4, I give some reasons why non-seriousness does not conflate with insincerity.
6 A similar definition had already been proposed by Searle (1975: 346).
7 Illocutionary point corresponds to the *essential condition* stated in Searle (1969: 63), which requires that the speaker intend her utterance as the performing of a specific (type of) act.

target group (it is not idiosyncratic), and (ii) it occurs just because the speaker is a member of that group (regardless of what she says and of how she says it). In what follows, I will concentrate on sexual refusals. The misrecognition will involve the illocutionary point of women's intended refusals, their status or position (which represents a crucial preparatory condition for refusing), and the sincerity of their words.

3. Four Types of Silencing

Before getting to the heart of the matter, let me address a preliminary question. As said above, locutionary acts are not the only kind of act we perform with words. With this in mind and following Langton (1993), we can enrich our intuitive notion of 'silencing' by analysing it along the speech levels introduced by Austin. The question to clarify is how someone's actions or words may silence the voice of someone else. First of all, one may literally shut someone else up by means of physical coercion (*e.g.*, gagging her), psychological violence (*e.g.*, threatening her), or institutional norms (*e.g.*, issuing legal restrictions on freedom of speech) (*locutionary silence*). Second, one may recognize but disregard someone's speech acts, so as to prevent her from obtaining her perlocutionary goals (*perlocutionary frustration*). Finally, one may deprive someone else's words of illocutionary potential by fostering a hostile communicative climate that interfere with the hearer's recognition of (certain components of) the force of the speaker's act, so that to cause its misfiring (*illocutionary disablement*). The silencing argument against pornography – as presented by H&L – involves the latter sense of silencing. When H&L argue that pornography silences women, they mean that it deprives women of illocutionary potential thereby preventing them from doing certain things with words – *e.g.*, refusing sexual advances. Pornography may silence women's illocutions through different mechanisms. It is time to examine them in detail, starting from the mechanism of uptake failure which gives place to what I label *essential silencing*.

3.1. Essential Silencing

Let's imagine the following scenario. A man approaches a woman for sex. She attempts to refuse it by saying "No". The man does not care and goes ahead forcing sex on her. According to H&L, the failure of the woman's refusal can be due to the missed acknowledgement by the man of her illocutionary intention. The key element here is *uptake* (*i.e.*, the hearer's recognition of the meaning and the force of the locution), which H&L treat as a necessary condition for illocution. If uptake is not achieved, then the woman's "No" cannot count as a refusal. Her act is unavoidably null, it does not take effect. How can pornography prevent women from achieving uptake? The answer provided is the following. Many pornographic stories include "favourable" rape depictions – that is, they represent reluctant women who give in to sexual pleasure upon being raped. Some presuppositions are obviously required for making sense of depictions like these. For instance, *women enjoy violent sex*, *women fantasize about rape*, *women's utterance of "No" is part of the game* (Langton & West 1999). The result is that some consumers of pornography may come to believe that, in saying "No", women *do not intend to refuse* – and this explains why women's acts are not recognized as they are meant to be taken.

Recasting H&L's proposal in Searlian terms, silencing can be conceived as a speech act failure constituted by the (systematic and unjust) misrecognition of the *illocutionary point* of the speaker's act. Due to some interfering factor, the addressee misses to acknowledge that the speaker's act meets its *essential condition*.

Before going any further, notice that H&L's view can be construed in (at least) two ways: (i) in some relevant contexts, a woman's "No" does not count as a refusal; (ii) in some relevant contexts, a woman's "No" counts as a consent. On the first reading, silencing originates from the failed recognition of the illocutionary point of the speaker's act. On the second, by contrast, it results from the hearer's ascription of a reversed illocutionary point to the act the

speaker is attempting to perform. On the one hand, pornography brings about *illocutionary disablement*; on the other, it brings about *illocutionary distortion* as well[8]. The latter reading has been proposed by Nellie Wieland (2007), who has suggested that, in H&L's account, pornography is *convention-setting* for it makes it the case that, in real-life sexual encounters, the expression "No" comes to mean *yes*. But if it is correct that pornography enacts such a linguistic convention, then men are right in interpreting women's "No" as consent moves. This gives rise to a highly undesired consequence: if the man has obtained his victim's consent, then he is no longer a rapist (Wieland 2007: 452-453). To avoid this problem, we may buy into the first of the above mentioned readings (Maitra & McGowan 2010). In so doing, however, we would be faced with a question: if the man's uptake of the woman's "No" is neither one of refusal nor one of consent, what kind of act does he attribute to her? The most plausible answer is *none*. In saying what she says, the woman is doing nothing but play-acting. If so, *essential silencing* comes to overlap with *seriousness silencing* (see Section 3.4).

3.2. Authority Silencing

McGowan (2009) identifies an alternative type of illocutionary silencing. On her account, refusals are *authoritative speech acts*: they require that the speaker have authority in the right domain. To see how authoritative speech works, imagine that a private tries to command a general to open fire against the enemy. Even if the general grasps the illocutionary point of the private's putative act (*i.e.*, he grasps his intention to command), the private nevertheless fails to illocute since he lacks the required authority. Sexual refusals, McGowan claims, are more similar to commands than to assertions as they can be performed only by speakers who have authority over their own bodies. Although every woman has this kind of authority simply in virtue of being a person, some men may fail to recognize it. Indeed, since the peculiar kind of pornography we are referring to presents women as mere tools to meet male sexual desire, its habitual consumers may come to regard them as having no authority whatsoever[9]. In depicting women in postures of sexual submission, pornography transmits the idea that women do not have rights over their own bodies, but men do – thereby making it impossible for them to exercise certain forms of practical authority[10].

Let's look at this type of silencing through the Searlian lens. As I have said, it involves the addressee's failure to ratify the speaker's authority. That is, while grasping the illocutionary point of the speaker's act, the recipient misrecognizes the obtaining of a crucial *preparatory condition* for refusals.

This form of silencing – as McGowan (2009: 493) acknowledges – does not appear to be a free speech issue. Whether or not it is the case turns on the more general question of what the right to free speech entails. I cannot fully address this question here, but, for my present aims, it may suffice to say that, in protecting freedom of speech, liberals (quite uncontroversially) want to protect speakers' freedom to communicate ideas and opinions to others – and the right to be free from systematic communicative interference seems to be constitutive of the "freedom to communicate ideas"[11]. On this view, *essential silencing* impinges on freedom of speech (the woman's "No" does not mean *no* to the hearer – that is, it fails to communicate the idea of refusal at all), while *authority silencing* does not, for it

8 The notion of 'illocutionary distortion' is borrowed from Green (2016), though I use it in a slightly different sense, meaning only those cases in which a speaker's utterance is given a reversed uptake.

9 McGowan's account develops some ideas outlined by Langton. See, for example, Langton (1993: 325: «A woman who prohibits sexual advances [...] has authority within the local domain of her own life, her own body [...]. If pornography prevents her from refusing, then pornography destroys her authority»).

10 Sbisà (2009) comes independently to analogous conclusions.

11 For a full discussion of what the right to free speech implies see West (2003).

involves no communicative breakdown (what the speaker says is indeed understood as it is meant to be taken).

3.3. Sincerity Silencing

Because of certain beliefs spread by pornography, when women refuse sex they might be systematically taken as insincere. The relevant beliefs may include *women always want sex* or *women tend to be coy so as not to appear promiscuous*. In order to account for this possibility (and developing an idea already put forth by Hornsby [1993: 42]), McGowan (2014) has introduced the notion of 'sincerity silencing'. In addition to illocutionary intentions concerning what kind of act one intends to perform, speakers also have sincerity intentions concerning the sincerity degree of their illocutions. Consequently, besides H&L silencing (*essential silencing*), there might be another type of silencing occurring when the addressee, while understanding the speaker's illocutionary intention, mistakenly believes that she is acting insincerely. Adopting Searle's framework, we could regard *sincerity silencing* as resulting from the hearer's misrecognition that the speaker's act meets the *sincerity condition*. Since in refusing sex the speaker expresses both a *desire* (to deter the hearer from having sex with her) and an *intention* (not to have sex with him), a sexual refusal is (sincerely) silenced when the hearer fails to take both the desire and the intention the speaker manifests with her "No" as psychological states she truly has. Notice that, in McGowan's (2014: 466) view, the audience's acknowledgment of the speaker's sincerity is necessary for successful communication. She argues for this thesis as follows. When a speaker performs a speech act, she usually pragmatically presupposes her own sincerity. If the recipient fails to grasp that presupposition, the speaker will succeed in communicating only part of what she is trying to get across. This means that *sincerity silencing* is a (partial) communicative failure, and thus constitutes a free speech violation. McGowan's argument appears to me not wholly convincing. Firstly, it is not clear why in performing an authoritative speech act – as refusals are – the speaker presupposes her own sincerity, but not her own authority (as McGowan herself implies by arguing that *sincerity silencing* is a violation of freedom of speech, whereas *authority silencing* is not). Secondly, partial communicative failure is, in a sense, unavoidable (it is extremely rare that addressees succeed in grasping *all* allusions or subtle implicatures of speakers' utterances), and this seems a good reason to deny its constituting a free speech infringement. McGowan (*ibid.*) is aware of this difficulty and tries to meet it by arguing that not everything a speaker intends to communicate is on a par. When the addressee fails to recognize nuances of allusions or implicatures, such a failure is not troubling. But, when a sexual refusal is mistakenly taken as insincere, the addressee's failure undermines the entire point of refusing (*i.e.*, to stop sexual advances). However, distinguishing between relevant and irrelevant implicit contents is not at all that easy, for the "degree of relevance" depends not only on the type of act performed but also on the context of utterance. Think to the act of refusal: when it is performed in the context of unwanted sex, the "sincerity presupposition" appears to be communicatively fundamental, but when it is performed in the context of, say, unwanted food, the same presupposition is much less important. In the light of these considerations, I do not regard *sincerity silencing* as a free speech impingement. *Sincerity silencing*, like *authority silencing*, does not involve any communicative failure: the woman's "No" does mean *no* to the hearer, even though it is systematically acknowledged as insincere.

3.4. Seriousness Silencing

According to Langton (1993: 321), the woman who is silenced in the sense of (what I have called) *essential silencing* is like the actor on the stage, who, after shouting for fire as part of the play, tries vainly to warn the audience of a real fire[12]. Unlike Langton, I prefer to save the distinction between the two accounts, since they are not equivalent. In the "essential" account, the hearer misses the illocutionary point of the speaker's act, but seems to recognize that she is illocuting. On the contrary, in the "play-acting" account, the hearer takes the speaker as performing a *non-serious* act – that is, as not illocuting. Moreover, although play-acting has something in common with insincerity (in both cases, the speaker is in a sense pretending), *seriousness silencing* does not conflate with *sincerity silencing*. While a non-sincere refusal (like a non-sincere promise) is still a refusal (or a promise), a play-acted refusal is not at all an illocutionary act, since it neither makes the speaker responsible for the performed act nor constitutes an attempt to carry out the perlocutionary object of the play-acted illocution[13]. Since speech act theorists consider seriousness as a background condition for illocution[14], we could regard *seriousness silencing* as resulting from the hearer's misrecognition that the speaker's utterance meets a fundamental *precondition* for illocuting[15]. Due to the rules pornography contributes to make valid in sexual contexts, men falsely believe that a woman's "No" is nothing but a line in a script, a move in the "game of sex" meant to increase her partner's arousal. Since the speaker's words are *stricto sensu* meaningless to the hearer, this form of silencing may properly amount to a free speech infringement.

Before folding our map, it is worth stressing that, although the above kinds of silencing involve only three of the seven components of the illocutionary force (*i.e.*, illocutionary point, preparatory conditions, and sincerity conditions), other forms of silencing may result from the misrecognition of the components of the force left over[16]. Consider Rebecca Kukla's (2014: 445-446) example of the female boss, who tells her male workers what to do but is very often disobeyed. The low obedience rate can be accounted for in different manners. One possible explanation is that, because of her gender, the workers take her utterances as *requests* rather than *orders*. According to Searle & Vanderveken (1985: 201), orders diverge from requests since they have a different mode of achievement. While requests leave the addressee free to refuse, to order somebody to do something is to direct her in a manner which does not leave that possibility open. If the boss' orders are systematically heard (and responded to) as requests, her disablement is due to the hearers' misrecognition of the mode of achievement of her acts. In spite of understanding the illocutionary point of the boss' utterances (*i.e.*, to get them to do something), the workers fail to recognize that her words do not give them the option of refusing; consequently, they feel free not to do what she tells them to do.

12 The example is from Davidson (1984: 269).

13 While Sbisà (2009: 353) takes the "play-acting" account as an instance of perlocutionary frustration, I see it as a proper form of illocutionary disablement (see Section 4 for discussion).

14 Austin (1962: 22); Searle (1969: 57).

15 Seriousness could also be understood as a preparatory condition for *every* illocutionary act. The dynamics producing *seriousness silencing*, so construed, would be analogous to those producing *authority silencing*.

16 This is one of the advantages of adopting Searle's conception of illocutionary force: it provides some insights in order to identify further types of silencing not yet considered in the debate on the phenomenon.

4. Failure, Defectiveness, and Misrecognition

As should be clear at this point, illocutionary silencing is a multifaceted phenomenon. By adopting Searle's framework, I have tried to integrate its diverse facets into a unique model. However, somebody may contend that not all types of silencing we have seen are instances of illocutionary disablement, since *some but not all* cause the speaker's act to misfire. In this section, I will try to see which kinds of silencing have the power to nullify the speaker's act and which kinds lack such a power.

According to Searle, every attempt to perform an illocutionary act can be (i) fully successful (when all conditions for success are met), (ii) successful though defective (when some non-necessary or additional conditions do not obtain), (iii) unsuccessful (when some necessary conditions are not satisfied). In both Austin's and Searle's theories, the successful securing of uptake figures among the necessary conditions for illocution[17]. It follows that *essential silencing* – which has been defined in terms of *uptake failure* – has the power to make the speaker's act null, and therefore falls fully into the category of 'illocutionary disablement'. Analogous remarks are relevant to *seriousness silencing*. In order to constitute some sort of illocutionary act an utterance has to be serious, and (crucially) it must be recognized as such. Why? Because the recognition of the speaker's seriousness bears on uptake: if the hearer fails to acknowledge the utterance as serious the whole act is given *no uptake at all*. There is an easy (but misguided) objection to my last statement, which could be put this way: *seriousness silencing* does involve some sort of uptake for the hearer understands the utterance as a *specific* fictional act – *e.g.*, as a fictional act of refusing sex (and not, say, consenting to it). This objection can be addressed by regarding fictional acts as *simulated speech acts*, which have a pretended force and achieve a fake uptake. Suppose that an actor on the stage, declaiming the last line of a monologue, says "There goes my life. Tonight I'll commit suicide". Though the audience understands the actor's utterance as having the (pretended) force of a prediction, nobody would ascribe him the performance of that illocution (even in the case, in declaiming those words, the actor meant to make a real prediction). *Because of the context*, the actor's words are given no true uptake. Similarly, although the man understands the woman's "No" as having the (pretended) force of a refusal, *because of the context* (and the rules that govern conversation in that context), her utterance does not secure any true uptake – and this is why her act ends up misfiring.

Authority and *sincerity silencing* are a somewhat different story. Let's start from *sincerity silencing*. As it is clear, the speaker's sincerity is not necessary for illocuting (it is always possible to perform an insincere speech act by expressing a psychological state one does not have), nor (*a fortiori*) is its recognition. Similar considerations apply to *authority silencing*. Even though the speaker's authority *is* a necessary condition for refusing, the hearer's recognition of that authority seems to be non-necessary or additional. (Since the man is wrong in regarding the woman as lacking the required authority, her "No" does count as a refusal)[18]. Apparently, *authority* and *sincerity silencing* fall into the category of 'perlocutionary frustration': the speaker whose authority or sincerity is denied is precluded from achieving *by* her words the intended effects on her audience. However, I take them as genuine instances of illocutionary disablement, in that they systematically (and unjustly) make the speaker's act unsuited even to *invite* the appropriate response on the part of the hearer. In the case of sexual refusals, the woman's "No" should *invite* the man to stop sexual advances (*n.b.*, this does

17 Austin (1962: 115-116); Searle (1969: 47).

18 The metaphysics of authority is a contentious issue. Somebody may claim, for instance, that having authority largely amounts to being recognized as having it and that authority is nothing but a social construction. On this way of thinking, if women are socially denied jurisdiction over their own bodies, then they do lack it. However, if at least some social agents (*e.g.*, women themselves) acknowledge women's authority to refuse sex, then gender prejudice cannot have the final say. I thank Tasneem Alsayyed for raising this point. See also Sbisà (2009: 355-356).

not imply that her "No" should *achieve* that effect)[19]. If he does not even consider that she has refused (that is, if the woman's utterance cannot invite its conventional response), the flaw is *illocutionary* (even if she succeeds in performing the act in question).
Despite their being more or less powerful, all the discussed kinds of silencing are forms of illocutionary disablement. If pornography silences women by undermining their ability to illocute rather than by merely causing their perlocutionary frustration, then it harms women in a particularly pernicious way. Perlocutionary acts (as opposed to illocutionary acts) are not under the speaker's full control, but rather related to the peculiar context of utterance as well as to the audience the speaker is addressing. A speaker may try to produce some perlocutionary effects without succeeding, or may not intend to achieve a certain sequel which occurs nonetheless. As troubling perlocutionary frustration is (especially when it gives rise to rape), it does not impair the very ability of the speaker to do things with words. Rather, it increases the degree of uncertainty implied in each and every performance of a perlocutionary act. On the contrary, the speaker whose words are divested of illocutionary potential is impeded in doing what – *under fair circumstances* – she would have been able to do simply by being heard as doing it. Illocutionary silencing causes the speaker's acts to be null or heavily defective without there being any reason for the failure but *the speaker is a member of a disempowered social group*, and thus constitutes an insidious form of discursive injustice[20].

5. Conclusion

In this paper I have tried to sketch a map of the main kinds of illocutionary silencing discussed in recent philosophical literature. By looking at the debate on the harms of pornography, I have identified four types of silencing (*essential*, *authority*, *sincerity*, and *seriousness silencing*), which in turn have been analysed in the frame of a unique model based on Searle's definition of 'illocutionary force' as a set of conditions for success. In the light of this definition, illocutionary silencing has been regarded as a speech act failure constituted by the systematic and unjust misrecognition on the part of the hearer of a specific component of the force of the speaker's act. It emerged from my analysis that neither *authority* nor *sincerity silencing* amounts to a free speech infringement, since they do not involve any communicative failure. Thus, they cannot be employed in support of MacKinnon's claim that pornography violates women's freedom of speech. In conclusion, I have argued that, despite reasonable hesitation, all the discussed kinds of silencing are genuine instances of illocutionary disablement, and thereby constitute pernicious forms of discursive injustice[21].

REFERENCES

Austin, J. L. (1962), *How to Do Things With Words*, Oxford University Press, London;
Davidson, D. (1984), "Communication and Convention", In Id., *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford, pp. 265-280;
Green, M. (2016), "Speech Acts: An Annotated Bibliography", *Oxford Bibliographies Online*, DOI: 10.1093/OBO/9780195396577-0300;
Hornsby, J. (1993), "Speech Acts and Pornography", *Women's Philosophy Review*, 10, pp. 38-45;

19 In Austin's (1962: 116) framework, the inviting of a certain response is an illocutionary effect, whereas the response itself is the speaker's perlocutionary goal, and the actual achievement of that goal constitutes the performance of the intended perlocutionary act.
20 Kukla (2014) uses the expression 'discursive injustice' to indicate a peculiar distortion of the path from speaking to uptake. Here I use it in a looser sense to refer to those cases in which the speaker's illocutions turn out to be flawed just because of her group membership. All types of silencing analysed above involve discursive injustice (in my loose sense).
21 I am grateful to Claudia Bianchi for detailed and helpful comments on an earlier version of this paper.

Hornsby, J. (1995), "Disempowered Speech", *Philosophical Topics*, Vol. 23, No. 2, pp. 127-147;
Hornsby, J. (2014), "Pornography and 'Speech'", In Coleman, L., and Held, J. M. (eds.), *The Philosophy of Pornography*, Rowman & Littlefield, London, pp. 129-146;
Hornsby, J., Langton, R. (1998), "Free Speech and Illocution", *Legal Theory*, 4, pp. 21-37;
Kukla, R. (2014), "Performative Force, Convention, and Discursive Injustice", *Hypatia*, Vol. 29, No. 2, pp. 440-457;
Langton, R. (1993), "Speech Acts and Unspeakable Acts", *Philosophy & Public Affairs*, Vol. 22, No. 4, pp. 292-330;
Langton, R., West, C. (1999), "Scorekeeping in a Pornographic Language Game", *Australasian Journal of Philosophy*, Vol. 77, No. 3, pp. 303-319;
MacKinnon, C. (1987), *Feminism Unmodified*, Harvard University Press, Cambridge;
MacKinnon, C. (1993), *Only Words*, Harvard University Press, Cambridge;
Maitra, I. (2012), "Subordinating Speech", In Maitra, I., and McGowan, M. K. (eds.), *Speech and Harm*, Oxford University Press, Oxford, pp. 94-120;
Maitra, I., McGowan, M. K. (2010), "On Silencing, Rape, and Responsibility", *Australasian Journal of Philosophy*, Vol. 88, No. 1, pp. 167-172;
McGowan, M. K. (2009), "On Silencing and Sexual Refusal", *The Journal of Political Philosophy*, Vol. 17, No. 4, pp. 487-494;
McGowan, M. K. (2014), "Sincerity Silencing", *Hypatia*, Vol. 29, No. 2, pp. 458-473;
Sbisà, M. (2009), "Illocution and Silencing", In Fraser, B., and Turner, K. (eds.), *Language in Life, and a Life in Language*, Emerald, Bradford, pp. 351-357;
Searle, J. R. (1969), *Speech Acts*, Cambridge University Press, Cambridge;
Searle, J. R. (1975), "A Taxonomy of Illocutionary Acts", In Gunderson, K. (ed.), *Minnesota Studies in the Philosophy of Science*, VII, University of Minnesota Press, Minneapolis, pp. 344-369;
Searle, J. R., Vanderveken, D. (1985), *Foundations of Illocutionary Logic*, Cambridge University Press, Cambridge;
West, C. (2003), "The Free Speech Argument Against Pornography", *Canadian Journal of Philosophy*, Vol. 33, No. 3, pp. 391-422;
Wieland, N. (2007), "Linguistic Authority and Convention in a Speech Act Analysis of Pornography", *Australasian Journal of Philosophy*, Vol. 85, No. 3, pp. 435-456.

SESSION

SESSION 4

EMOTIONS AND INTERSUBJECTIVITY – TYPICAL DEVELOPMENT AND PATHOLOGIES

THOMAS FUCHS
University of Heidelberg
thomas.fuchs@med.uni-heidelberg.de

INTERCORPOREALITY AND INTERAFFECTIVITY[1]

abstract

According to phenomenological and enactive approaches, human sociality does not start from isolated individuals, but from intercorporeality and interaffectivity. To elaborate this concept, the paper introduces (1) a concept of embodied affectivity, regarding emotions as a circular interaction of the embodied subject and the situation with its affective affordances. (2) This leads to a concept of embodied interaffectivity as a process of coordinated interaction, bodily resonance, and 'mutual incorporation' which provides the basis for a primary empathy. (3) Finally, developmental accounts point out that these empathic capacities are also based on an intercorporeal memory that is acquired in early childhood.

keywords

empathy, intercorporeality, interaffectivity, bodily resonance, intercorporeal memory

Phenomenology and Mind, n. 11 - 2016, pp. 194-209
DOI: 10.13128/Phe_Mi-20119
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

Since the rise of cognitive psychology in the early 1970s, 'social cognition' has become the dominant concept in social psychology and cognitive neuroscience to denote the processes of social understanding and empathy. It is commonly based on a representationalist point of view: Internal cognitive mechanisms such as a 'theory of mind' enable an observer to 'mentalise' or 'mind-read', i.e. to infer others' hidden states of mind. Regardless of whether these mechanisms are described as akin to a scientific 'theory' or rather as a mental 'simulation' routine (Premack and Woodruff 1978, Baron-Cohen *et al.* 1985, Stich and Nichols 1991, Carruthers and Smith 1996), the general framework has mostly remained true to its origins in classical cognitivism and representationalism. This corresponds to the currently predominant concepts of *emotions:* they are not regarded as embodied responses to meaningful situations, being perceivable in the bodily expression and conduct of another person, but rather as internal cognitive appraisals of environmental stimuli (Solomon 1976, Lyons 1980, Nussbaum 2001). Since emotions are thus in principle hidden to others, empathy can only be based on mind-reading or on simulating others' mental states inside oneself.

However, in most everyday situations, we don't use any introspection, simulation routines or inferences when we interact with others. Instead, we immediately perceive their intentions and emotions in their expressive behaviour as being related to a meaningful context (Gallagher 2008). As I will show in the following, it is mainly *bodily resonance* which conveys an intuitive understanding of others' emotions in our embodied engagement with them. The ongoing interaction induces, on a pre-reflective level, a process of mutual modification of bodily and emotional states, thus enabling a primary form of empathy without requiring any representations. Accordingly, phenomenological and enactive approaches to sociality do not start from isolated individuals and their respective inner states, but from a priority of *intercorporeality* and *interaffectivity* (Gallagher 2001, Fuchs & De Jaegher 2009, Froese & Fuchs 2012, De Jaegher 2015).

To elaborate this concept, I will introduce (1) a general concept of *embodied affectivity*: it conceives emotions not as inner mental states residing within individuals (even less their brains), but as encompassing spatial phenomena that connect the embodied subject and the

1 This is a slightly revised version of my chapter „Intercorporeality and Interaffectivity", in: C. Meyer, J., Streeck, J. S., Jordan (Eds.) *Intercorporeality: Emerging Socialities in Interaction.* Oxford: Oxford University Press, 2017.

situation with its affective affordances in a circular interaction. (2) This leads to a concept of *embodied interaffectivity:* in every face-to-face encounter, the partners' subject-bodies are intertwined in a process of bodily resonance, coordinated interaction and 'mutual incorporation' which provides the basis for an intuitive empathic understanding. It can also give rise to self-sustaining interaction patterns that go beyond the behavioural dispositions of isolated individuals. According to this concept, emotions may not primarily be localized within a single individual, but should rather be conceived as phenomena of a shared intercorporeal space in which the interacting partners are involved.
(3) Finally, developmental accounts of intersubjectivity point out that sharing and understanding each other's feelings is also based on an *intercorporeal memory* or *implicit relational knowledge* that is acquired in early childhood. It conveys a basic sense of social attunement or a 'social musicality'. Primary empathy as mediated by embodied interaction may subsequently be extended by higher-level cognitive capacities such as perspective-taking and imaginary transposition. Nevertheless, intercorporeality and interaffectivity remain the basis of social cognition.

1. Embodied affectivity

To begin with, we should abandon the idea that emotions are only 'mental' phenomena, and the world is bare of any affective qualities. The introjection of feelings into an inner 'psyche' is a heritage of Platonic and, later on, Cartesian dualism. In fact, we do not live in a merely physical world; the experienced space around us is always charged with affective qualities. We feel, for example, the hilarity of a party, the sadness of a funeral march, the icy climate of a conference, the awe-inspiring aura of an old cathedral or the uncanniness of a sombre wood at night. Such atmospheric effects are evoked by physiognomic or expressive qualities of objects as well as by intermodal features of perception such as rhythm, intensity, dynamics, etc.[2] Emotions no less emerge from situations, persons and objects which have their expressive qualities, and which attract or repel us. The peculiar intentionality of emotions (see e.g. Solomon 1976, Frijda 1994, De Sousa 2010) relates to what is particularly *valuable and relevant* for the subject. In a sense, emotions are ways of perceiving, namely attending to salient features of a situation, giving them a significance and weight they would not have without the emotion. Referring to Gibson's (1979) concept of affordances (that means, offerings in the environment that are available to animals, such as a tree being 'climbable', water 'drinkable', etc.), one could also speak of *affective affordances*: things appear to us as 'interesting', 'expressive', 'attractive', 'repulsive', 'uncanny', and so on.
How do we experience the affective qualities or affordances of a given situation? Emotions are experienced through what I call *bodily resonance* (Fuchs 2000, 2013a). This includes all kinds of local or general bodily sensations: feelings of warmth or coldness, tickling or shivering, pain, tension or relaxation, constriction or expansion, sinking, tumbling or lifting, etc. There is no emotion without at least the slightest bodily sensations and movement tendencies. Of course, when I am moved by an emotion, I may not always be aware of my body; yet being afraid, for instance, is not possible without feeling a bodily tension or trembling, a beating of the heart or a shortness of breath, and a tendency to withdraw. It is *through* these sensations that we are anxiously directed towards a frightening situation, even if we do not notice them. Therefore, bodily feelings and action tendencies should not be conceived as a mere by-product or add-on, distinct from the emotion as such, but as the *very medium* of affective intentionality. The body

2 These structural qualities of perception have been particularly explored by Gestalt psychology (cf. Koehler 1992) and nowadays been rediscovered by infant research as so-called „vitality affects" (Stern 1985). See also the phenomenology of affectivity in Fuchs 2000: 193-217.

is a 'resonance body', a most sensitive 'sounding board' in which every emotion reverberates (James 1884).

This leads to an *embodied and extended conception of emotions* (Fuchs 2013a):

Emotions emerge as specific forms of a subject's bodily directedness toward the values and affective affordances of a given situation. They encompass subject and situation and therefore may not be localized in the interior of persons (be it their psyche or their brain). Rather, the affected subject is engaged with an environment that itself has affect-like qualities. For example, in shame, an embarrassing situation and the dismissive gazes of others are experienced as a painful bodily affection which is the way the subject *feels* the sudden devaluation in others' eyes. The emotion of shame is extended over the feeling person and his body as well as the situation as a whole.

Emotions further imply two components of bodily resonance:

- *a centripetal or affective component,* i.e. being affected, 'moved' or 'touched' by an event through various forms of bodily sensations (e.g. the blushing and 'burning' of shame);
- *a centrifugal or 'emotive' component,* that means, a bodily action readiness, implying specific tendencies of movement (e.g. hiding, avoiding the other's gaze, 'sinking into the floor' from shame). Other tendencies are approach (desire, anger), avoidance (fear), being-with (enjoyment, confidence), rejection (disgust), dominance (pride) or submission (humility, resignation) (cf. Frijda 1986). Thus, in emotions "we are moved to move toward or against or away" (Sheets-Johnstone 1999: 267).

Taken together, emotions may be regarded as *circular interactions or feedback cycles* between the embodied subject and the situation it is confronted with (cf. fig. 1): Being affected by the affective affordances or value features of the situation *('affection', impression)* triggers a specific bodily resonance which in turn influences the emotional perception of the situation *and* implies a corresponding expression and action readiness *('e-motion').* Embodied affectivity consists in the whole interactive cycle which is crucially mediated by the resonance of the feeling body.

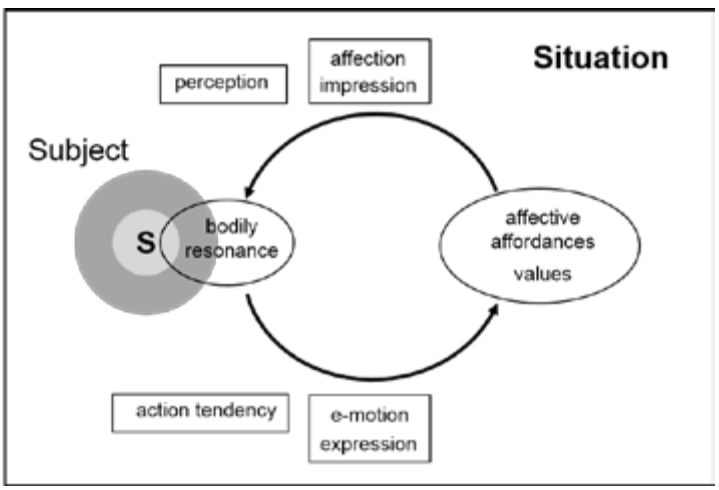


Fig. 1: Embodied affectivity [3]

3 Adapted from Fuchs 2013a: 623.

Bodily resonance thus acts as the medium of our affective engagement in a situation. It imbues, taints and permeates the perception of this situation without necessarily stepping into the foreground. In Polanyi's terms, bodily resonance is the *proximal*, and the perceived situation is the *distal,* component of affective intentionality, with the proximal or tacit component receding from awareness in favour of the distal (Polanyi 1967). This may be compared to the sense of touch which is at the same time a self-feeling of the body ('proximal') and a feeling of the touched surface ('distal'); or to the subliminal experience of thirst ('proximal') which first becomes conspicuous not as such, but as the perceptual salience of water flowing nearby ('distal'). Similarly, bodily resonance initially goes unnoticed; only at a certain level of intensity, it becomes conscious as such (for example when one's hear starts to pound in fear).

2. Embodied interaffectivity

If we now turn to the social sphere, we can see that the cycle of 'affection' and 'emotion', impression and expression involves another person as a specific 'affective affordance'. Emotions thus become *interactive phenomena* which are not only felt from the inside, but also displayed and visible in expression and behaviour, often as bodily tokens or rudiments of action.[4] The facial, gestural and postural expression of a feeling is part of the bodily resonance that feeds back into the feeling itself, but also induces processes of *interaffectivity:* Our body is affected by the other's expression, and we experience the kinetics and intensity of his emotions through our own bodily kinaesthesia and sensation. Our body schemas and feelings expand and 'incorporate' the perceived body of the other. This creates a dynamic interplay which forms the basis of social understanding and empathy, and which I will describe as *mutual incorporation* (Schmitz 1989, Leder 1990, Fuchs & De Jaegher 2009).

Incorporation is a pervasive characteristic of the 'lived' or subjective body *(Leib)* which always transcends itself and connects with the environment. This is the case for example in every skilful handling of instruments, as when playing piano and letting the fingers find their way by themselves; or when a blind man probes his environment with a stick and feels the surface at the top of it. In such cases, the instrument is integrated into the body motor schema like an extension of the body, subjectively felt as 'melting' or being at one with the instrument. However, such incorporation also occurs with other people, even at a distance. An example for this is fascination. Thus, we may listen to a spellbinder, hanging on his every word, and feel being drawn towards him. Or we may watch the performance of an acrobat on a high wire with a mixture of fascination and anxiety. Our lived body extends and connects with the acrobat's swinging movements; we may even be prompted to co-movements.

In a more subtle and subliminal way, such coupling occurs in every face-to-face encounter: Two cycles of embodied affectivity (fig. 1) become intertwined, thus continuously modifying each partner's affective affordances and bodily resonance, as illustrated in fig. 2:

Let us assume that A is a person whose emotion, e.g. anger, manifests itself in typical bodily (facial, gestural, interoceptive, adrenergic, circulatory, etc.) changes. His pre-reflectively experienced or lived body thus functions as a felt 'resonance board' for his emotion: A feels the anger as the tension in his face, as the sharpness of his voice, the arousal in his body etc. These proprio- and interoceptive bodily feelings may be termed *intra-bodily resonance.* Now this

4 According to Darwin (1872), emotional expressions once served particular action functions (e.g. baring one's teeth in anger to prepare for attack), but now accompany emotions in rudimentary ways in order to communicate these emotions to others. Evolutionary psychologists have advanced the hypothesis that hominids have evolved both with increasingly differentiated facial expressions and with sophisticated capabilities of understanding these affect displays. In any case, though strongly varying between and within cultures, emotional expression is a crucial facet of interpersonal communication in all societies.

resonance also implies an *expression* of the emotion, that means, the anger becomes visible and is perceived by A's partner B. Moreover, the expression will also produce an *impression*, namely by triggering corresponding or complementary bodily feelings in B. Thus, A's sinister gaze, the sharpness of his voice or expansive bodily movements might induce in B an unpleasant tension or even a jerk, a tendency to withdraw, etc. Thus, B not only sees the anger immediately in A's face and gestures, but also senses it with his own body, through his own intra-bodily resonance.

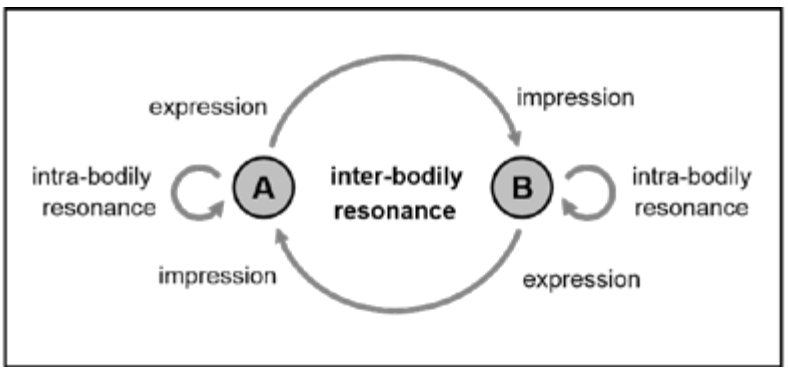


Fig. 2: Mutual incorporation and inter-bodily resonance[5]

However, it does not stay like this, for the impression and bodily reaction caused in B in turn becomes an expression for A; it will immediately affect his bodily reaction, change his expression, however slightly, and so forth. This creates a circular interplay of expressions and reactions running in split seconds and constantly modifying each partner's bodily state, in a process that becomes highly autonomous and is not controlled by the partners. They have become parts of a dynamic sensorimotor and inter-affective system that connects both bodies by reciprocal movements and reactions, that means, by *inter-bodily resonance*.[6] Of course, the signals and reactions involved proceed far too quickly to stand out discretely and become conscious as such. Instead, both partners will experience a specific feeling of being bodily connected with the other in a way that may be termed *mutual incorporation:* Each lived body with its sensorimotor body schema reaches out, as it were, to be extended by the other. This is accompanied by a holistic impression of the interaction partner and his current state (for example his anger), and by a feeling for the overall atmosphere of the shared situation (for example a tense atmosphere).[7]

No 'mental representation' or 'mind reading' is implied in this process. There is no strict or dualistic separation between an inside and an outside at all, as if a hidden mental state in A

5 Adapted from Fuchs 1996.

6 The term ‚resonance' is not infrequently used in social neuroscience as underlying social cognition or empathy (e.g., ‚motor resonance', Gallese 2001, ‚affective resonance', Decety & Chaminade 2003, Decety & Meyer 2008), but usually related to mirror neuron mechanisms or shared neural representations, without being spelled out phenomenologically. This is being done here. For a phenomenological account including references to neural resonance mechanisms, see also Gallagher 2012.

7 As the example shows, this interbodily resonance and resulting atmosphere is by no means restricted to 'harmonious' social situations – interpersonal conflicts may create most intense forms of mutual resonance and atmospheric tensions. Even situations of withdrawal, exclusion or ostracizing are felt as an inter-bodily field of negative field forces, so-to-speak (cf. Fuchs 2007).

produced certain external signs that B would have to decipher.[8] For A's anger may not be separated from its bodily resonance and expression; and conversely, B does not perceive A's body as a mere object, but as a living, animate and expressive body that he is in contact with. This expressiveness is concentrated and intensified in the *gaze*: the other is for me not somewhere 'behind his gaze', but he is visible in it. Even more, I experience the other's gaze *as being seen by him*, or in other words*, I see him seeing me (seeing him)*. The irrefutable evidence of the other's embodied presence emerges from being mutually affected by each other, or from inter-affection.

Nor may the process be described as a bodily *simulation* of the other's state that we project back onto him. We certainly do not simulate another's angry gaze or voice, even less his anger, in order to notice it, but we rather feel tense, threatened or invaded by his expressive bodily behaviour. That means, our resonance is *complementary* and does not mirror the other's expression. But even in case of *corresponding* resonance (e.g., the other's smile inducing a smile in me), there is no need for the complicated mechanism of an 'as-if'-simulation and back projection. Instead, my own intra-bodily resonance is simply *implied in my perception* of the other, namely as its 'proximal', tacit component. Bodily sensations, tensions, action tendencies, etc. that arise in our interaction do not serve as a separate simulation of the other, but are part and parcel of our embodied mutual perception.

Using the phenomenological distinction between the *subjective, lived body* (*Leib)* and the *physical, living body* (*Körper)*, we can also describe this mutual intertwinement as follows: The *lived body's impression* in the one person (A) becomes a *living body's visible expression* for the other person (B), and vice versa: the impression produced in B's lived body becomes a living body's expression for A. Thus, it is the peculiar 'chiasmatic'[9] structure of the body as the turning point of interior and exterior, as both *Leib* and *Körper*, which enables the interlacement of self and other in the process of mutual affection and perception. This analysis may be regarded as an articulation of Merleau-Ponty's concept of 'intercorporeality' (*intercorporéité,* Merleau-Ponty 1960), by which he intended to complement Husserl's account of intersubjectivity as the constitution or 'appresentation' of the other by a conscious ego (Husserl 1960).[10] Intercorporeality means a pre-reflective intertwining of lived and living bodies, in which my own is affected by the other's body as much as his by mine, leading to an embodied communication:

> The communication or comprehension of gestures comes about through the reciprocity of my intentions and the gestures of others, of my gestures and the intentions

8 Of course, humans are also able to control their emotional expressions, that means, to withhold, to enact or to feign emotions to a certain extent, for example when playacting, lying or cheating (such as in poker games), etc. However, this does not mean that emotions are actually inner or disembodied states:

(1) The success of those artificial or feigned expressions depends precisely on the primary interaffectivity: If we would not normally perceive others' expressions *as* their emotional states, we couldn't easily be deceived either. Spontaneous bodily resonance remains the default mode of non-verbal understanding.

(2) The control of one's emotional bodily resonance is restricted to the movement of the voluntary muscles, whereas the autonomous bodily resonance and general action readiness remains outside of control. This may be demonstrated by measuring muscle tonus, heart rate, skin conductance (think of a lie detector), etc. Hence, emotions are still embodied states, even though we may inhibit their motor expression to a certain extent.

9 This term was used by Merleau-Ponty in his later work to denote that our embodied subjectivity may not be located merely in either the body's touching or in its tangibility, in our interior or exterior, but in their intertwining, or where the two lines of a chiasm intersect without coinciding (Merleau-Ponty 1968).

10 Certainly Husserl's concept of intersubjectivity may not be restricted to the *Cartesian Meditations* in which he developed the idea of appresentation. Other accounts, which in part anticipate Merleau-Ponty's concept of intercorporeality, are given in *Ideen II* and *Experience and Judgement* (Husserl 1952, 1973).

> discernible in the conduct of other people. It is as if the other person's intentions inhabited my body and mine his (Merleau-Ponty 1962).

As we can see, intercorporeality leads to the opposite of a representationalist account: Primary social understanding is not an inner modelling in a detached observer, but the other's body extends onto my own, and my own extends onto the other. There is no component within the interactive cycle that represents ('stands for') another one, for this would require it to be separated from the cycle, thus reconstructing inside what is discernible outside. But in intercorporeality, inside and outside are not separate domains, but only directions of motion within an ongoing mutual transition between *expression* and *impression*, or 'e-motion' and 'affection'.

To illustrate this once more, let us imagine a football play in which one player sees his teammate raise his arms rejoicing at a goal. According to representationalism, there are objects 'out there', in this case a body, whose features are transmitted to the retina, then further processed by the brain in order to create an internal representation of the other's body, which is then combined with a theory of mind or simulation mechanism, resulting in the appraisal: "he is happy". Instead of this linear concept, an intercorporeal approach emphasizes the circular sensorimotor dynamics within the dyad of embodied agents. Both partners are linked to form an encompassing system through mutual perceptions and reactions. Grasping, pointing, handing-over, moving-towards, smiling, crying, etc. – all these are not just external behaviours that we have to furnish with a meaning by way of inference, but through our bodily resonance, they become inherently expressive and meaningful actions. Thus, the footballer will immediately perceive the other as an 'affective affordance', so to speak, and empathically sharing his pleasure, he might also perceive him as a person he could hug. His understanding is interactive from the start, and might easily result in spontaneously embracing his teammate. No simulation or introspection is necessary to share the pleasure – the embrace is merely the manifestation of both player's intercorporeality and interaffectivity.

3. The development of intercorporeality

The concepts of embodied affectivity and interaffectivity describe universal structures of (inter-)subjective experience. On the other hand, there is no doubt that the forms of emotional expression and interaction are to a large extent shaped by one's individual biography as well as by one's cultural background. From early childhood on, patterns of interaction with others are sedimented in the infant's implicit or bodily memory, resulting in what may be called *intercorporeal memory* (Fuchs 2008, 2012). A similar concept is the *habitus* as a culture- or class-specific set of dispositions, skills, styles, tastes and demeanour, which are adopted through everyday interactions according to one's social and cultural background (Bourdieu 1990). Apart from these acquired dispositions and habits, there is also a history of particular interactions between two partners, resulting in what may be called an encompassing or *dyadic body memory*. In this last section, I want to give a short account of these diachronic aspects of intercorporeality.

3.1. Intercorporeal memory

The notion of *body memory* may be used to denote all forms of implicit memory that are mediated by the body and actualised without explicit intention or recollection in our everyday conduct (Fuchs 2008, 2012). It thus comprises all those customs, habits, manners, and practices which are performed unreflectively or 'as a matter of course'. It is a procedural memory for motion patterns such as walking or dancing, for handling instruments such as a bicycle or a keyboard, for familiar *gestalts* of perception, for complex spatial situations (for example finding one's bearings in a dwelling or a town), and last not least for the habitual bodily interactions with others. Thus, body memory conveys the familiarity with recurrent typical

situations and enables skilled interactions with the social environment. Whereas explicit recollection is directed to the past, implicit or body memory re-enacts the past through the body's present performance; in other words, it may be regarded as our 'lived past'.
Implicit or body memory does not mean a subpersonal motion program realised by a body machine without a subject. When I am dancing, the rhythmic movements originate from my body without a need to steer them deliberately – and yet I am living in my movements, I sense them in advance, and I can modulate them according to the rhythm that I feel: I myself am dancing, and not a ghost in a body machine. The movements of my body are at my disposal, I am aware of my capacities, and thus I feel up to my present task as an embodied being. In the last analysis, all capacities acquired in life are integrated into a primordial capability of the embodied subject, a basic sense of agency or 'I can' (Husserl 1952: 253).
More specifically, *intercorporeal memory* means a pre-reflective, practical knowledge of how to interact with others in face-to-face encounters which is acquired already in early childhood. This is enabled by implicit or procedural learning which the infant is capable of from birth on, whereas explicit and autobiographical memory only develop in the course of further brain maturation, i.e. from the 2nd year of life on (Welzer & Markowitsch 2005, Bauer 2006). Let us look at these early learning processes in more detail:
Infants are attuned from birth to social interactions, in particular by showing a heightened attention to faces and their expressions (Valenza *et al.* 1996; Turati *et al.* 2002). Research studies conducted during the last two decades have mostly found that they are also able to imitate adults' gestures like sticking out their tongue, opening their mouth, frowning and others (Meltzoff and Moore 1977, 1989). Thus it seems that the newborn's body schema is characterized by an intermodal openness that immediately allows it to transpose the perceived expressions into its own proprioception and movement.[11] Since bodily imitation evokes corresponding feelings as well, we may assume that a mutual bodily resonance also emerges in the early dyadic interactions. This assumption is supported by the fact that 6-8 weeks olds already engage in proto-conversation with their mothers by smiling and vocalizing (Trevarthen 1979, 1993). They both exhibit a finely tuned coordination of movements, rhythmic synchrony and mirroring of expressions, that has often been compared to a couple dance (Gopnik & Meltzoff 1997: 131). Infant and caregiver also follow a turn-taking pattern, shifting the roles of agent and recipient in a non-random sequence (Jasnow and Feldstein 1986).[12]
As early as in the first months, infants become capable of discerning emotions such as happiness, sadness, and surprise in the postures, movements, facial expressions, gestures, and vocalizations of others (Hobson 2005: 39ff.). The basis for this is that different sense modalities can have the same 'kinematics' and thus express the same affect. Infants both perceive and express affects as the intermodal extract of rhythmic, melodic, vocal, facial and gestural characteristics: *'crescendo'* or *'decrescendo', 'accelerando'* or *'ritardando'*, rising or falling movements, flowing or explosive dynamics, etc. Thus, for example, the feeling of joy and the

11 Recent research with larger samples and a wider range of gestures presented to the infants challenges these results, finding no significant excess of matching over non-matching reactions (Oostenbroek *et al.* 2016). But even if it turns out that imitation is not an innate capacity, but learnt in the course of mutual exchanges during the first months, it still functions as a major component of what Trevarthen (1979) has termed "primary intersubjectivity".
12 It should be noted, however, that early communication does not ideally mean "complete synchronisation", but always includes sequences of matches, mismatches and subsequent "repair" which are also important for the infant to experience a difference between the mother as different from him- or herself, as shown my micro-analyses of the interaction (Tronick and Cohn 1989, Beebe *et al.* 1997). Synchronisation thus means a rhythmic or phasic harmonisation, not complete congruence.

various expressions of joy have similar intermodal dynamics, and this is the basis for the direct perception of others' emotional states even in earliest childhood.
According to Stern (1985), the temporal flow patterns and kinematics of the interaction that are felt by both partners result in *affective attunement,* which may be regarded as equivalent to inter-bodily resonance and mutual incorporation. This means that emotions are not primarily observed in others, but they emerge as *dyadic affective states*, often as an intense pleasure or joy (Tronick 1998). The shared affect during a joyful playing situation between mother and infant may not be divided and distributed among them. It arises from the 'between', or from the situation in which both are immersed. Thus, affects are not enclosed in an inner mental sphere to be deciphered from outside, but emerge, change and circulate between self and other in the intercorporeal dialogue.
The recursive patterns of these interactions are now sedimented in the infant's body memory, leading to what Lyons-Ruth and Stern (1998) have called *implicit relational knowing*. This means a pre-reflective knowledge or skill of how to deal with others – how to share pleasure, elicit attention, express happiness, avoid rejection, re-establish contact etc. The infant acquires specific interactive schemes („schemes of being-with", Stern 1985) and body micro-practices that are needed for a growing range of interactions. Implicit relational knowing is a temporally organised, 'musical' memory for the rhythm and dynamics that are subliminally present in the interaction with others (Stern 1985, Amini *et al.* 1996). It also implies an *interaffective memory* for the specific 'feel' of the shared vitality contours and the emotions that they carry.
Already in the first few months of life, infants demonstrate a memory for shared interaction sequences through the way they expect their mothers to react. Babies quickly learn to which emotional expressions parents respond, are spurned to action, or rather dismiss, etc. An impressive demonstration of the emergence of intercorporeal memory is enabled by the 'still face' experiment (Weinberg & Tronick 1996): during play with her infant, the mother is asked to assume a blank facial expression and stare straight ahead for two minutes. Babies from 2-3 months on usually react with clear irritation and unease – the expected resonance from the mother is absent – and they try in every possible way, with gestures and vocalizing, to elicit the mother's attention and to re-establish the familiar form of contact. More specifically, two groups of children can be differentiated (Field 1984):

1) Infants of sensitive and lively mothers remain active even in the still-face situation, and obviously expect to bring their mother back into contact in this way.
2) Infants of mothers who are more detached and lacking resonance (for instance, because of post-partum depression), react differently; in the beginning they are restless, but quickly become passive and helpless. In other words, they have not learned to effectively use their behaviour to induce contact. Later on these children show a marked weakness in attachment (Field *et al.* 1988).

These considerations and research results may suffice to show that intercorporeality is shaped by individual experiences which to a large extent date back to early childhood interactions, but have a lasting and even lifelong influence on our embodied relationships with others.

3.2. The concept of habitus

From a sociological point of view, body memory, and in particular intercorporeal memory, may also be seen as the carrier of the *habitus,* that means, a set of dispositions, skills, styles, tastes and behaviour that is shared by the members of a community, class or culture. The habitus is acquired by the individual through practical immersion in the life world, in the course of interactive experiences, mimetic learning, implicit routines, rituals etc. As Bourdieu notes (1977: 94), societies treat "... the body as memory; they entrust to it in abbreviated and practical, i.e. mnemonic, form the fundamental principles of the arbitrary concepts of culture. The principles embodied in this way are placed beyond the grasp of consciousness." This

embodied memory is realised by an "implicit pedagogy" (l.c.); it does not require purposeful instructions, deliberate imitation or other kinds of explicit learning. On the contrary, the habitus becomes a second nature which effectively guides one's behaviour, all the more as it is not conscious *as* a habitus.

This is also important for our concepts of social understanding: The homogeneity of the habitus as the shared body memory of a community or culture entails that the embodied practices are immediately evident or foreseeable on the background of a given situation. This provides a primary, non-inferential understanding of others without requiring conscious transposition, perspective-taking or simulation ("what would I do in his place?"). Growing up and being immersed in a common practical context results in tacit knowledge of the "rules of the game" and of typical interactive sequences (Condon 1979). Like the players of a football team on the field, the members of a culture normally understand each other intuitively, anticipate the next moves and know how to react, without a need to resort to deliberation, to a theory of mind or mentalising procedures.

The habitus thus becomes the basis of common sense or *sensus communis*, namely as a practical sense of embodied social customs and interactions which constitutes the pre-reflective background of social life. In this way, collective styles of intercorporeality and interaction are passed on from one generation to the next without becoming explicit – in an unconscious, collective history. Thus Merleau-Ponty's notion of intercorporeality gains an additional aspect: It means not only the primary familiarity of our bodies with each other, or their pre-reflective communication, but also the entanglement of human bodies in a shared history that is preserved in their implicit collective memory.

When relating the phenomenological concept of body memory to Bourdieu's habitus, however, one precaution seems necessary. Bourdieu himself has repeatedly criticised phenomenological approaches as relying merely on subjective descriptions, thus failing to account for the sociological determination of commonsensical practices (cf. Bourdieu 1977: 3f., 183; 2000: 132). While this may be disputed in particular in the light of Husserl's later work on the habitus (see e.g. Husserl 1973: 122) and of Schütz' social phenomenology of the life world (Schütz 1962), there is a reverse tendency in Bourdieu to adopt an objectivist and structuralist point of view which results in an overly deterministic picture of human sociality. One of the main reasons for this is that he mistakenly regards all non-representational forms of knowledge and habit ('knowing how') as necessarily 'non-conscious' and as such not accessible to conscious modification or control. However, as I have pointed out above, the realisation of bodily skills and practices is not a subpersonal process but open to modifications or even explicit changes by the subject. In principle, persons are always capable of acquiring new social skills; they may even reflect on their incorporated social background and then deliberately strive to transform their own habitus – though this may take a great deal of time and effort. Thus, the concept of body memory does not involve a deterministic stance but is open to individual creativity and social change.

3.3. Dyadic Body Memory

As we have seen, each body forms an extract of its past history of experiences with others that are laid out in its intercorporeal memory. As a result of learning processes, which are in principle comparable to the acquiring of motor skills, social agents shape and enact their relationships according to the patterns they have extracted from earlier and earliest experience. On the other hand, each particular interaction, when repeated, also acquires its own history, thus pre-figuring and constraining future interactions between the respective partners. One may, for example, develop a specific style of interacting with a close friend which is only possible with this person and re-emerges immediately even after years of separation. Then the intercorporeal memories of both partners unite to form a *joint procedural*

field that suggests and preordains typical interactions and shared experiences. It may also be regarded as a space of two body schemas which are attuned to each other through sensorimotor patterns generated by the shared history: rituals of welcoming, repertoires of gestures, postures, movements, voice pitch and even dialects which one 'falls into' in the presence of the other. Hence, we may say that there is, in a certain sense, a memory of the interactive process itself, or a *joint or dyadic body memory*.

Let us take another example, namely a well-trained couple of dancers who move easily with the rhythm of the music, make the suitable turns even on the slightest signals of the other, and whose hands and bodies find each other without guidance of the gaze. Both partners apply their procedural and intercorporeal skills, and yet they move and feel in a way that is only possible within the interaction. Together they create the spatio-temporal pattern of the dance which in turn draws them into its superordinate dynamics, implying the mutual incorporation already described above. Their kinaesthetic body schemes literally extend and connect with each other to form an overarching dynamic process. Rhythm and melody particularly support this incorporation by providing dynamic constraints for the movements of both partners. Each of them behaves and experiences differently from how they would do outside of the process (or with another partner); their interaction has gained an autonomy of its own (De Jaegher & Di Paolo 2007).

Where shall we localize this memory of joint dancing and other skilful bodily interactions? On the one hand, the superordinate or dyadic bodily system certainly has no natural substrate for forming a memory – it only consists of the present connection of two lived bodies in which the respective dispositions have formed. A 'memory of the system' must finally be based on the individual and biological memories of the agents involved in order to become effective for behaviour. On the other hand, it is only together that the individuals are in a position to re-enact the dyadic pattern, which justifies to attribute this memory or knowing-how in a sense to the dyad itself. It is a memory not in the sense of 'I can' but of *'we can'.* As we can see, intercorporeality means more than inter-bodily resonance and mutual incorporation: it may also be regarded as an overarching system which over time gains its own pattern, autonomous dynamics and peculiar history.

4. Conclusion

I have outlined a concept of primary or pre-reflective intersubjectivity which is based on embodied affectivity and interaffectivity. It conceives emotions not as inner mental states that have to be deciphered or inferred from external cues, but as expressive, dynamic forces which affect individuals through bodily resonance and connect them with one another in circular interactions. In face-to-face encounters, each partner's lived body reaches towards the other to form an overarching system through inter-bodily resonance and mutual incorporation. According to this concept, social understanding is primarily based on intercorporeality; it emerges from the interactive practice and coordination of embodied agents. We do not need to form internal models or representations of others in order to understand and communicate with them; as bodily subjects, we are always already involved in a shared affective and expressive space.

In social contacts, our lived bodies become extended such that they are intertwined with those of others in a way that prevents any conceptual or ontological reduction to isolated entities. This applies both to current interactions and to the history of interactive patterns. From early childhood on, social understanding and empathy develop as a practical sense, a musicality for the rhythms, dynamics and patterns of interactions with others. Intermodal kinematics and bodily resonance are key to attuning and sharing each other's affects within the primary dyad. Thus, in a non-mentalising way, infants are already able to perceive the emotions and intentions in the actions of others, in their postures, gestures and facial expressions, as related

to the context of the common situation. This provides a primary understanding without recourse to a concept of mental states.

Moreover, developmental research indicates that empathy is based on an intercorporeal memory or an implicit relational knowledge of how to interact with others that is acquired in early childhood and conveys a basic sense of social attunement. In each social encounter, both partners unconsciously re-enact a history of embodied socialization and relationships that have shaped their styles of interacting, their empathic skills and intuitions as well as their class- and culture-specific habitus. Finally, embodied intersubjectivity can also give rise to self-sustaining interaction patterns that go beyond the behavioral dispositions of isolated individuals. They may be attributed to a memory of the intercorporeal system and its partially autonomous dynamics, an 'attractor landscape', so-to-speak, that is actualised and modified in every new encounter of the participating agents.

To be sure, this embodied and enactive concept does not exhaust the possibilities of empathic understanding and intersubjectivity. On the basis of primary bodily empathy, we are also able to explicitly represent, to imagine or to question the other's situation. This happens in particular when their behaviour seems ambiguous, or when an irritation or misunderstanding occurs. Through additional information and inference, we can then try to enhance our understanding, infer possible hidden intentions and in this way often deepen our empathy. A further possibility is to transpose oneself into the other's situation and imagine how one would feel or react in his place. Here we use a kind of simulation which I prefer to call 'imaginary transposition' or 'perspective taking' (Fuchs 2013b).

Such higher-level forms of social understanding develop later in life, mainly from the 2nd to 4th year. Knowledge about others that is based on language and narrative reports plays a crucial role for these later stages of intersubjectivity (Gallagher & Hutto 2008, Hutto 2009). However, "... it is only gradually and with considerable input from adults that children eventually come to conceive of something like 'bodies' on the one hand and 'minds' on the other" (Hobson 1993: 117). Even then one may argue that the concept of 'Theory of Mind' was misleading from the start, since the very term presupposes a kind of inferential and scientific approach to others' allegedly hidden minds as the standard mode of intersubjectivity. Yet the need for such an approach arises mainly in situations in which intercorporeal and verbal communication becomes ambiguous or extremely restricted, such as poker games or the Cuba crisis. Sophisticated, detective-like cognitive capacities are neither necessary nor sufficient to enable empathic intersubjective relations. Despite those later developments, our everyday social understanding remains based on embodied intersubjectivity.

REFERENCES

Amini, F., Lewis, T., Lannon, R., Louie, A., Baumbacher, G., McGuiness, T., and Schiff, E. Z. (1996), "Affect, attachment, memory: contributions toward psychobiologic integration", *Psychiatry* 59, pp. 213-239;

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985), "Does the autistic child have a "theory of mind"?" *Cognition* 21, pp. 37–46;

Bauer, P. J. (2006), "Constructing a past in infancy: A neuro-developmental account", *Trends in Cognitive Sciences* 10, pp. 175-181;

Beebe, B., Lachmann, F., and Jaffe, J. (1997), "Mother—infant interaction structures and presymbolic self- and object representations", *Psychoanalytic dialogues* 7, pp. 133-182;

Bourdieu, P. (1977), *Outline of a Theory of Practice*, Translated by Richard Nice, Cambridge University Press, Cambridge, UK;

Bourdieu, P. (1990), "Structures, habitus, practices", In P. Bourdieu, *The logic of practice*, Stanford University Press, Stanford, CA, pp. 52-79;

Bourdieu, P. (2000), *Pascalian Meditations,* Stanford University Press, Stanford, CA;
Carruthers, P., and Smith, P. K. (eds.) (1996), *Theories of theories of mind*. Cambridge University Press, Cambridge;
Condon, W. S. (1979), "Neonatal entrainment and enculturation", In Bullowa, M. (Ed.), *Before Speech*, Cambridge University Press, Cambridge, pp. 131-148;
Darwin, C. (1872/1904), *The expression of emotions in man and animals*, Murray, London;
Decety, J., and Chaminade, T. (2003), "Neural correlates of feeling sympathy", *Neuropsychologia* 41, pp. 127-138;
Decety, J., and Meyer, M. (2008), "From emotion resonance to empathic understanding: A social developmental neuroscience account", *Development and psychopathology* 20, pp. 1053-1080;
De Jaegher, H. (2015), "How we affect each other: Michel Henry's 'pathos-with' and the enactive approach to intersubjectivity", *Journal of Consciousness Studies* 22, pp. 112-132;
De Jaegher, H., and Di Paolo, E. A. (2007), "Participatory sense-making: An enactive approach to social cognition", *Phenomenology and the Cognitive Sciences* 6, pp. 485-507;
De Sousa, R. (2010), *Emotion,* Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/archives/spr2010/entries/emotion.
Field, T. (1984), "Early interactions between infants and their postpatum depressed mothers" *Infant Behavior and Development* 18, pp. 1-3;
Field, T., Healy, B., Goldstein, S., Perry, S., Bendell, D., Schanberg, S., Zimmermann, E. A., and Kuhn, C. (1988), "Infants of depressed mothers show "depressed" behavior even with nondepressed adults", *Child Dev* 59, pp. 1569-1579;
Frijda, N. H. (1986), *The emotions*, Cambridge University Press, Cambridge;
Frijda, N. H. (1994), "Varieties of affect: Emotions and episodes, moods, and sentiments", In Ekman, P., and Davidson, R. J. (eds.), *The nature of emotion: Fundamental questions*, Oxford University Press, New York, pp. 59-67;
Froese, T., and Fuchs, T. (2012), "The extended body: A case study in the neurophenomenology of social interaction", *Phenomenology and the Cognitive Sciences* 11, pp. 205-236;
Fuchs, T. (1996), "Leibliche Kommunikation und ihre Störungen", *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie* 44, pp. 415-428;
Fuchs, T. (2000). *Leib, Raum, Person. Entwurf einer phänomenologischen Anthropologie,* Klett-Cotta, Stuttgart;
Fuchs, T. (2007), "Psychotherapy of the lived space. A phenomenological and ecological concept", *American Journal of Psychotherapy* 61, pp. 432-439;
Fuchs, T. (2008), "Das Gedächtnis des Leibes", In Fuchs, T., *Leib und Lebenswelt. Neue philosophisch-psychiatrische Essays*, Graue Edition, Kusterdingen, pp. 37-64;
Fuchs, T. (2012), "The phenomenology of body memory" In Koch, S., Fuchs, T., and Müller, C. (eds.) *Body memory, Metaphor and Movement*, John Benjamins, Amsterdam, pp. 9-22;
Fuchs, T. (2013a), "The phenomenology of affectivity", In Fulford, B., Davies, M., Graham, G., Sadler, J., and Stanghellini, G. (eds.), *The Oxford Handbook of Philosophy and Psychiatry,* Oxford University Press, Oxford, pp. 612-631;
Fuchs, T. (2013b), "The phenomenology and development of social perspectives", *Phenomenology and the Cognitive Sciences* 12, pp. 655-683;
Fuchs, T., and De Jaegher, H. (2009), "Enactive intersubjectivity: Participatory sense-making and mutual incorporation", *Phenomenology and the Cognitive Sciences* 8, pp. 465-486;
Gallagher, S. (2001), "The practice of mind: theory, simulation or primary interaction?", *Journal of Consciousness Studies* 8, pp. 83-108;
Gallagher, S. (2008), "Direct perception in the intersubjective context", *Consciousness and Cognition* 17, pp. 535-543;
Gallagher, S., Hutto, D. (2008), "Understanding others through primary interaction and

narrative practice" In Zlatev, J., Racine, T. P., Sinha, C., and Itkonen, E. (eds.), *The shared mind: perspectives on intersubjectivity*, John Benjamins, Amsterdam, pp. 17-38;
Gallagher, S. (2012), "Neurons, neonates and narrative: From embodied resonance to empathic understanding", In Foolen, A., Lüdtke, U. M., Racine, T. P., and Zlatev, J. (eds.) *Moving ourselves, moving others: Motion and emotion in intersubjectivity, consciousness and language*, John Benjamins, Amsterdam, pp. 167-196;
Gallese, V. (2001), "The 'shared manifold' hypothesis. From mirror neurons to empathy", *Journal of Consciousness Studies* 8, pp. 33-50;
Gibson, J. (1979), *The ecological approach to visual perception,* Houghton Mifflin, Boston;
Gopnik, A., and Meltzoff, A. N. (1997), *Words, Thoughts, and Theories*, MIT Press, Cambridge, MA;
Hobson, R. P. (1993), *Autism and the development of mind,* Lawrence Erlbaum Associates, Hillsdale NJ;
Hobson, R. P. (2002), *The cradle of thought*, Macmillan, London;
Husserl, E. (1952), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. II. Phänomenologische Untersuchungen zur Konstitution. Vol. IV, Husserliana* [Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Studies in the phenomenology of constitution], Martinus Nijhoff, Den Haag;
Husserl, E. (1960), *Cartesian meditations*, Transl. D. Cairns, M. Nijhoff, The Hague;
Husserl, E. (1973), *Experience and Judgment,* Northwestern University Press, Evanston;
Hutto, D. (2009), "Folk psychology as narrative practice", *Journal of Consciousness Studies* 16, pp. 9-39;
James, W. (1884), "What is an emotion?" *Mind* 9, pp. 188-205;
Jasnow, M., and Feldstein, S. (1986), "Adult-like temporal characteristics of mother-infant vocal interactions", *Child Development* 57, pp. 754-761;
Koehler, W. (1992), *Gestalt psychology: An introduction to new concepts in modern psychology,* Liveright, New York;
Leder, D. (1990), *The absent body,* University of Chicago Press, Chicago;
Lyons, W. (1980), *Emotion,* Cambridge University Press, Cambridge;
Lyons-Ruth, K., Bruschweiler-Stern, N., Harrison, A. M., Morgan, A. C., Nahum, J. P., Sander, L., Stern, D. N., and Tronick, E. Z. (1998), "Implicit relational knowing: its role in development and psychoanalytic treatment", *Infant Mental Health Journal* 19, pp. 282-289;
Meltzoff, A., and Moore, M. K. (1977), "Imitation of facial and manual gestures by human neonates" *Science* 198, pp. 75-78;
Meltzoff, A., and Moore, M. K. (1989), "Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms", *Developmental Psychology* 25, pp. 954-962;
Merleau-Ponty, M. (1960), "Le philosophe et son ombre", In *Signes,* Éditions Gallimard, Paris;
Merleau-Ponty, M. (1962), *The phenomenology of perception*, Humanities Press, New York;
Merleau-Ponty, M., Lefort, C. (1968), *The Visible and the Invisible,* Northwestern University Press, Evanston;
Nussbaum, M. C. (2001), *Upheavals of thought. The intelligence of emotions,* Cambridge University Press, Cambridge UK;
Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., and Slaughter, V. (2016), "Comprehensive longitudinal study challenges the existence of neonatal imitation in humans" *Current Biology* 26, pp. 1334–1338;
Polanyi, M. (1967), *The tacit dimension*, Anchor Books, Garden City, New York;
Premack, D., and Woodruff, G. (1978), "Does the chimpanzee have a theory of mind?", *Behavioral and Brain Sciences*, 1, pp. 515–526;
Schmitz, H. (1989), "Über leibliche Kommunikation", In H. Schmitz, *Leib und Gefühl. Materialien zu einer philosophischen Therapeutik*, Junfermann, Paderborn, pp. 175-218;

Schütz, A. (1962), *Collected Papers, Volume 1: The Problem of Social Reality*, Nijhoff, The Hague;
Sheets-Johnstone, M. (1999), "Emotion and movement. A beginning empirical-phenomenological analysis of their relationship", *J. Consciousness Studies* 6, pp. 259-277;
Solomon, R. (1976), *The passions*, Anchor/Doubleday, New York;
Stern, D. N. (1985), *The interpersonal world of the infant: A View from Psychoanalysis and Developmental Psychology*, Basic Books, New York;
Stich, S., and Nichols, S. (1991), *Folk psychology: Simulation or tacit theory*, Rutgers University Press, New Brunswick, NJ;
Trevarthen, C. (1979), "Communication and cooperation in early infancy: a description of primary intersubjectivity", In Bullowa, M. (ed.), *Before Speech*, Cambridge University Press, Cambridge, pp. 321-347;
Trevarthen, C. (1993), "The self born in intersubjectivity", In Neisser, U. (ed.), *The Perceived Self: Ecological and Interpersonal Sources of Self-knowledge*, Cambridge University Press, Cambridge, pp. 121-173;
Tronick, E. Z., and Cohn, J. F. (1989), "Infant-Mother Face-to-Face Interaction: Age and Gender Differences in Coordination and the Occurrence of miscoordination", *Child Development* 60, pp. 85-92;
Tronick, E. Z. (1998), "Dyadically expanded states of consciousness and the process of therapeutic change", *Infant Mental Health Journal* 19, pp. 290-299;
Turati, C., Simion, F., Milani, I., and Umiltà, C. (2002), "Newborns' Preference for Faces: What is Crucial?", *Developmental Psychology* 38, pp. 875–882;
Valenza, E., Simion, F., Cassia, V. M., and Umiltà, C. (1996), "Face Preference at Birth", *Journal of Experimental Psychology: Human Perception and Performance* 22, pp. 892–903;
Weinberg, M. K., and Tronick, E. Z. (1996), "Infant affective reactions to the resumption of maternal interaction after still-face", *Child Development* 67, pp. 905-914;
Welzer, H., and Markowitsch, H. (2005), "Towards a bio-psycho-social model of autobiographical memory", *Memory* 13, pp. 63-78.

CESARE MAFFEI
Vita-Salute San Raffaele University
maffei.cesare@hsr.it

VIOLA FUSI
Vita-Salute San Raffaele University
vio.fusi@gmail.com

EMOTION DYSREGULATION IN BORDERLINE PERSONALITY DISORDER: A LITERATURE REVIEW

abstract

The goal of this literature review is to describe the state of art related to emotion dysregulation in BPD and to illustrate a possible descriptive model.
As of today, no consensus in literature is reached, but, from a conceptual point of view, most of the authors agrees with Linehan's theory in understanding emotion dysregulation both as affective instability and as a frequent recourse to dysfunctional regulation strategies. The latter is what emotion dysregulation is from an operational point of view. The descriptive model explained in this paper is a possible way to bridge the conceptual and operative views of emotion dysregulation in BPD.

keywords

emotion dysregulation, BPD, affective instability, dysfunctional strategies, descriptive model

Phenomenology and Mind, n. 11 - 2016, pp. 210-220
DOI: 10.13128/Phe_Mi-20120
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

The *goal* of this review is to describe the state of art related to emotion dysregulation in BPD, trying to group first empirical and theoretical studies with experimental and neuro-cognitive papers that debate this issue and to group then those with similar theoretical and conceptual directions.
We reviewed the most relevant studies from 2003 to 2014. The studies taken into consideration are experimental and neuro-experimental ones together with theoretical reviews and papers.

1. Three theoretical macro-conceptualizations derived from literature review

Emotion dysregulation is acknowledged by most clinical models as the major and central clinical feature both in the pathogenesis and treatment of BPD, considered as *"primarily a disorder of the emotion regulation system"* (Linehan 1993: 43) and the main cause of additional emotional symptoms manifested (Glenn & Klonsky 2009; Putnam & Silk 2005).
Currently, a lack of conceptual clarity prevents to define unambiguously and precisely this construct:
in fact, we could identify f three possible macro-theoretical conceptualizations in the literature review.

1.1. The First Macro-Conceptualization: An Operative Definition of Emotion Dysregulation as Maladaptive and Inefficient Regulatory Strategies

This conceptualization associates to emotion dysregulation the meaning of a broad use of *maladaptive and inefficient behaviors or strategies* undertaken by individuals to manage their emotional experiences.
Even if disadaptive in most cases, these actions represent a regulatory attempt by BPD subjects to cope with the emotions experienced and not just being overwhelmed by them.
Despite the great number of classifications of regulatory strategies, the literature considers as reference model the *Modal Model of Emotion Regulation*.
As *Figure 1* shows, there are five prototypical strategies (Gross & Thompson 2007).
The first one is the s*ituation selection*, consisting in approaching or avoiding certain people, places, or activities in order to limit as much as possible the experience of negative emotions. *Situation modification*, refers to acting on specific situation-stimulus to modify its emotional impact. *Attentional deployment* is a strategy the individual chooses to focus on a specific aspect of certain situation.
The fourth strategy of *cognitive change* refers to the selection of possible meanings attributable to a specific situation in order to decrease its emotional impact.
The last strategy, *response modulation,* refers to the modulation, as direct as possible, of the emotional response from a physiological or experiential perspective.

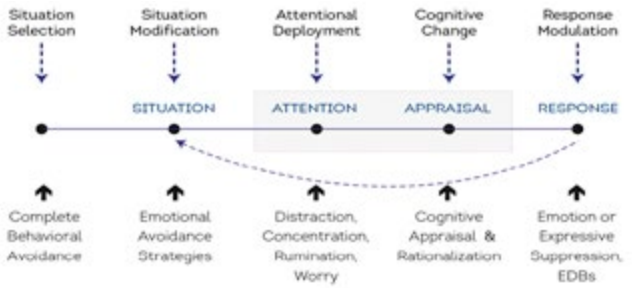


Fig. 1. The Modal Model of Emotion Regulation by Gross. *Above:* adaptive version; *below:* disadaptive version. Adapted from Fairholme *et al.,* 2010.

The difference between these strategies is the time they need in order to have a primary impact on the generative emotional process (Gross *et al.* 2006).
According to the above, the first four are "antecedent-focused" strategies, antecedent the activation of an emotional response, aimed to alter directional trajectory if this is disadaptive and not conform to individual's objectives.
Instead, the fifth and last strategy, response modulation, is a "response-focused" strategy, subsequent to the activation of the emotional response, aimed at regulating the current emotional state.
Emotion regulation strategies act more effectively as operational unconscious mechanisms: the frequency of their employment and possibly their success in past emotional situations are two determiners of the unconscious operation (Bargh & Williams 2007).
Within psychopathology, it is possible that the *presence* of a *maladaptive* emotion regulation strategy could be more deleterious than the *absence* of an *adaptive* one (Aldao *et al.* 2009).
Each of the five strategies in the Modal Model (Gross 2007), has its maladaptive version. The maladaptive form of *situation selection* is the *complete behavioral avoidance*, which is utilized by individuals to avoid situations of emotional distress: this strategy seems to give emotional relief in the short term but not in the long term.
Distraction, concentration, rumination and worry are all examples of a maladaptive *attention deployment: rumination* is a verbal-linguistic strategy that implies a cognitive focalization on previous negative situations or on their perceived failure (Fairholme *et al.* 2010). It is the most harmful of all maladaptive versions of attention deployment.
Cognitive appraisal and rationalization are disadaptive versions of the strategy of *cognitive change.* The first one denotes the formulation of realistic interpretations of the emotional situation. Differently, rationalization implies justificatory interpretations given by the individual in order to alter the current emotional situation.
Finally, *emotional or expressive suppression and EDBs* are disadaptive versions of *response modulation.* The term "suppression" means, behaviorally, a disguise of emotions that consequently are experienced through artificial behaviors (Gross & Thompson 2007) while, emotionally, indicates the inhibition of the emotional experience itself.
EDBs mean "Emotion-driven behaviors" that are specific actions engaged according to the relevance of the emotional experience itself.
Having explained the theoretical framework of the first conceptualization of emotion dysregulation, we can now analyze the studies.
According to Schmahl *et al.* (2014), emotion dysregulation refers to "the *frequency"* with which patients use maladaptive strategies to regulate emotions and it is a *driving force behind*

several serious dysfunctional behavioral patterns, including suicidal ideation, suicide attempts, maladaptive interpersonal behaviors and impulsive coping behaviors, such as alcohol abuse" (Schmahl *et al.* 2014: 2).

Other authors focalized their studies on *several aspects* of the emotion regulation strategies. For example, in their review Putnam & Silk (2005), showed that BPD subjects, *in the short term,* seem to be able to utilize efficient strategies to alter the emotional tenor of a stimulus, intensifying the valence and arousal of negative stimuli as well as prolonging the duration of the effect of the stimulus.

Therefore, although these are baneful attempts, individuals with BPD demonstrate an operative capacity to modify, somehow, the emotional quality of a stimulus.

Conversely, *in the long term,* these regulatory strategies become inconsistent with the goals of the individual and deleterious to interpersonal functioning, probably because BPD subjects tend to experience much more negative emotions over the time than positive or neutral ones.

Interesting findings by Pietrek *et al.* (2012) showed that at a neuromagnetic level, dysfunctional emotion regulation occurs at the level of response regulation, not at the level of input processing.

Other studies focalize their attention specifically on the *different types of dysfunctional emotion regulation strategies.*

For example, Dixon-Gordon (2014) examined in a sample of 84 BPD subjects, the role of positive and negative emotion differentiation in daily life in predicting urges for a range of maladaptive behaviors.

Emotion differentiation is an adaptive emotion regulation strategy, which describes the ability to make fine-grained distinctions between similarly valenced emotional states.

Not all the individuals have the same emotion differentiation capacities: some of them are high differentiators, who are able to distinguish between emotional states with similar valence (e.g. sadness, anger). On the contrary, others are low differentiators, because tend to describe their emotions in more global terms, based only on the valence property (e.g. pleasantness vs. unpleasantness), losing possible important information about their emotional experiences (Zaki *et al.* 2013).

The results showed that positive emotion differentiation is a potential protective factor in the relation between high BPD features and maladaptive behaviors; while negative emotion differentiation does not impact urges for these type of behaviors, suggesting that reducing impact urges may not be the mechanism by which emotion differentiation facilitates non-engagement in maladaptive behaviors.

Other studies analyzed *suppression*, a specific and recurrent dysfunctional emotion regulation strategy utilized by BPD subjects. Beblo *et al.* (2013), for example, investigated attempts to suppress or accept negative and positive emotions both in BPD patients and in healthy participants. Different important conclusions can be drawn from these findings.

First, the strategy of suppression is not always efficient in decreasing emotional negative states: rather, its action can paradoxically emphasize the intensity and occurrence of negative emotions. For this reason, the suppression strategy has been called "*ironic process*" because, regarding its initial purpose, it leads to an exactly opposite result (Lowenstein 2007).

Moreover, according to the authors, intense negative emotions may lead to fear of emotional arousal. In accordance with this assumption, through stimulus generalization this fear may expand to fear of positive emotions, which subsequently lead BPD subjects to suppress them. This process could explain why BPD individuals suppress emotions with positive valence.

Nevertheless, the attempt to suppress negative and positive emotions shows that these individuals *are not just acting out emotions without trying to regulate them*, confirming once again the operative definition of emotion dysregulation.

Last, but not least, the study of Vine & Aldao (2014) showed that *deficits in emotional clarity* might underpin regulatory dysfunctions and associated borderline symptomatology.
In addition, they stressed other two facets of impaired emotion regulation in BPD individuals: the *access to emotion regulation strategies* and, to a lesser degree, *attention-shifting ability.*
Chapman *et al.* (2011), focused particularly on the latter, showing that deficits in emotion regulation among persons with BPD specifically involve difficulty engaging response-focused emotion regulation strategies and *difficulty deploying attention away* from emotionally evocative stimuli and maladaptive attempts to control intense emotions.

1.2. The Second Macro-Conceptualization: Emotion Dysregulation Equated to "Affective Instability"

Differently from the above mentioned studies, which consider emotion dysregulation in terms of inefficient strategies engaged by BPD subjects to regulate intense felt emotions, other authors refer to this construct with the term of "*affective instability*", defined as marked intensity, reactivity and variability of moods (APA 2000). It is important to underline that this definition of emotion dysregulation leaves out every operative attempt (even if unsuccessful) made by individuals trying to manage their intense emotional state: in this case, it's the quality of the emotion itself that leads to an unstable and dysfunctional emotional arousal.
As reported by Doll *et al.* (2013), BPD is "characterized by "*stable instability*" of emotions, impulsivity, social relationships and self-image.
Some authors focalized on *emotional reactivity,* one of the component of affective instability. For example, Kuo *et al.* (2014) in their study underlined that *emotional reactivity is a key process in BPD*.
Moreover, they added that "though not always explicitly defined "emotional reactivity", many other models of BPD reference extreme changes in *emotional intensity or emotional lability* as key features of the disorder, including attachment-based, developmental, interpersonal, genetic and psychodynamic models" (*Ibidem:* 155).
The authors examined the differences in emotional reactivity in response to standardized and idiographic stimuli, and across three specific emotions (sadness, fear and anger). The findings show that idiographic or personally-relevant stimuli (primarily related to anger and sadness) were more effective in eliciting emotional reactivity in BPD compared to standardized stimuli. Another finding is that sufficient reactivity seem to be elicited *prior to engagement of regulation strategies* so that, emotion dysregulation would take place already at the level of the former, not necessarily with the presence of the latter.
Sansone & Sansone (2010) also examined emotional hyper-reactivity in BPD, seen as a low threshold for responsiveness and a greater responsiveness to the environment. Reviewing findings from clinical experience and empirical studies, they found out that BPD subjects may over-react to negative stimuli as well as positive or neutral ones. Their hyper-responsiveness in different environmental situations is related to relationship issues, mainly regarding themes of loss, abandonment and/or the encountering of limits.

1.3. The Third Macro-Conceptualization: Emotion Dysregulation as the Combination of Affective Instability and Maladaptive Emotion Regulation Strategies

The third and last definition could be seen as the product of the previous ones, because it explains emotion dysregulation as the *combination* of *affective instability,* including the *characteristics of the emotional response itself* (e.g. duration, intensity, frequency etc.), and the *maladaptive behaviors or strategies* engaged to manage the emotions experienced.
Mennin & Fresco (2010) explained these two domains of emotion dysregulation. The first relates to the *generative characteristics of the emotional experience*, including intensity, valence and durability of the emotional response. The second domain of dysregulation is related to the *regulative processes of an emotional experience*. Among these, there are *limited emotional knowledge and reduced awareness* that imply the individual's inability to clarify, classify and differentiate emotions from underlying motivations. These dysfunctions prevent to get a meaning from these experiences and respond effectively to the current situation; individuals could also have *cognitive negative reactions* to emotions, that imply an activation of negative expectations and beliefs regarding emotional states.
Another dysfunctional emotional process is the maladaptive management of responses to emotions: this feature is particularly characteristic of BPD and is indicated as a difficulty to recognize how and when increase or decrease responses to a specific current emotion. Consequently, individuals are not able to engage themselves in resolute behaviors or efficient regulatory strategies.
Dysregulation could occur in one of these features, such as for emotional intensity, which is a risk factor for the development of Borderline Personality Disorder.
The conceptual framework used for this conceptualization of emotion dysregulation is Linehan's biosocial theory (1993), emphasizes a biological emotional vulnerability that leads also to problems with emotion dysregulation in terms of an engagement of maladaptive strategies and a lack or limited access to more adaptive ones. In their theoretical review, Carpenter & Trull (2013) showed that emotion dysregulation is not an end-state, but a process incorporating multiple interactive components, derived from Linehan's biosocial theory. These components are an emotion sensitivity, a higher negative affectivity, a deficit of appropriate regulation strategies and a surplus of maladaptive regulation strategies. Emotion dysregulation seems to operate in a circular way, so that inadequate regulation strategies, derived from higher and unstable negative affectivity, reinforce in turn vigilance toward negatively valenced stimuli in the environment. Anyway, more research is needed to better elucidate the interactive way of operating of these four components.
Also Newhill *et al.* (2012) focalize their conceptualization of emotion dysregulation on Linehan's theory, but they took in consideration only the aspect of affective instability and not the operative disadaptive engaged strategies. In this case, emotion dysregulation consists of three dimensions: a high sensitivity or vulnerability to emotional stimuli, a high amplitude of emotional response to such stimuli and a slow return to emotional baseline after affective arousal.
Moreover, in their neuro-experimental study, Donegan *et al.* (2003) assessed amygdala reactivity in 15 BPD subjects compared to 15 NC subjects. They observed a greater left amygdala activation to facial expressions in BPD group compared with NC group, and a difficulty disambiguating neutral or threatening faces in the former too.
These findings discovered substantial amygdala activation as a key component of emotional vulnerability in BPD patients, especially in the contest of disturbed interpersonal relations and in processing emotional stimuli and reactions.
Besides in this neuro-scientific context, Ruocco *et al.* (2012), evaluated whether the magnitude of the volume reductions of hippocampus and amygdala and their associations with state-of-illness factors and psychiatric comorbid disorders could be considered as potential endophenotypes. The authors, reviewing 11 research studies, came to the conclusion that the

modest volume reductions of the amygdala and hippocampus cannot be attributed to illness state or comorbid psychopathology but they may hold promise as candidate endophenotypes for BPD.
Some authors add other causes/predictors to emotion dysregulation compared to those explicated by Linehan in the biosocial model.
For example Gorska (2013) added as causes of emotional dysregulation also the referential processes that indicate a dissociation of internal emotional schemas, all factors that seem to be conditioned more socially by relationships. The referential activity in BPD individuals stops at sub-symbolical level, so for the basic function of symbolization, mentalization or metacognition is then seriously compromised, leading to immature emotional representations, full of idiosyncratic details.
Consequently, for these subjects is impossible to integrate emotional experiences, the fundamental element that would have made emotion regulation possible.
If the previous studies refer to, eventually amplifying, Linehan's biosocial theory, others only partially support this model.
For example, Kuo & Linehan (2009) analyzed in particular three dimensions of the biosocial model (biological vulnerability, high emotional intensity and high reactivity) in a sample composed by 20 BPD, 20 SAD (*e.g. social anxiety disorder)* and 20 NC participants. Their findings revealed that if emotion dysregulation in BPD subjects is accounted for high baseline emotional intensity and biological vulnerability, this is not the case of high reactivity. In fact, these individuals seem to be not more reactive than nonclinical and socially anxious subjects. Moreover, it remains unclear whether emotional intensity is indeed a sequence of vulnerability (as proposed by Linehan's model) or whether it is an outcome of the transaction between an emotionally vulnerable individual and an invalidating environment.
Reeves *et al.* (2010) tested the main corollary of the biosocial theory, according to which emotional vulnerability and invalidation lead to emotional dysregulation, which ultimately leads to BPD symptoms.
In a very large nonclinical sample they found out that, contrary to predictions, a history of emotional invalidation was not related to self-reported BPD symptoms, suggesting that this variable may not play such an important role as predicted by Linehan's biosocial theory.
This finding was further corroborated by Gill & Warburton (2014), who showed in their study that emotionally validating parenting acts as a protective factor against the development of borderline traits, but it *doesn't correlate with emotional dysregulation.*
This finding is important because, according to the authors, it highlights that not all borderline etiology is mediated by emotional dysregulation. Other types of poor parenting, beyond the emotionally one, may play a role in BPD symptomatology.
Selby & Joiner (2009) proposed the Emotional Cascade Model, which is incremental to Linehan's theory.
In this conceptual framework, BPD arises from a complex network of interacting factors and at the center there are the emotional cascades. They originate primarily from an intense rumination process of negative events that results in emotional intensity and, consequently, in behavioral dysregulation.
In this case, *the strategy of rumination,* the disadaptive version of the strategy of attentional deployment (Gross & Thompson 2007), is the first responsible for both affective instability and behavioral dysregulation.
Moreover, according to this model, only dysregulated behaviors, in form of *distraction,* induce temporary reduction of negative emotion, leading to relief.
Finally, Beblo *et al.* (2010), investigated whether emotional intelligence was compromised in BPD subjects such that it could compromise the engagement of adaptive emotion regulation

strategies. Their findings showed that intense emotions could trigger dysfunctional emotion regulation strategies even when patients possess sufficient theoretical knowledge about optimal regulation strategies. Consequently, it seems plausible that an individual's selection of emotion regulation strategies is more influenced by the current emotional state rather then (or not exclusively) by abstract knowledge.

2. Discussion

This systematic review reassumes the state of art related to emotion dysregulation in BPD as evidenced by literature. The state of art on emotion dysregulation in BPD underlines a great diversity of theoretical and conceptual lines derived mainly from the research results. From a detailed analysis of the literature, we can state that the emotion dysregulation is a core feature of BPD, but it has not yet been precisely defined the way it occurs within the disorder itself. From a conceptual point of view, most of the authors, taking in consideration the theory of Linehan, agrees in understanding emotion dysregulation both as affective instability relating to the quality and intensity of emotions, and as a frequent recourse to regulatory strategies in managing maladaptive emotions. However, from the operational point of view, according to some authors, emotion dysregulation would manifest within the BPD only as use of maladaptive strategies such as rumination and suppression, which would strengthen the intensity of the current emotional state. Further studies are needed to achieve greater consistency among the various theoretical positions and to test empirically the ways in which the emotional dysregulation occurs within the disorder.

3. Conclusions. Emotion Dysregulation in Borderline Personality disorder: a possible descriptive model

A possible model of what happens in the "here and now" to BPD subjects could start from the interaction among values, objectives and action. Values are abstract concepts, they are socially shared, and the realisation of them is what could give worthiness to life. Objectives are both abstract and concrete, that is contextually determined. Objectives are the embodiment of values in real life through action, that is strictly contextual. Actions happen in the "here and now" and choosing an action means excluding other actions. If reality is probabilistic before acting, it becomes deterministic when we are acting.

To explain the link between the behavioural dysfunctionality of BPD subjects to their difficulty in emotion regulation we could start from two possible options.

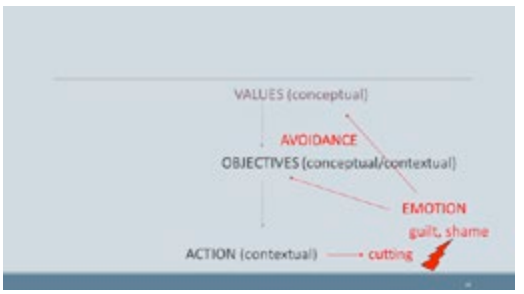
According to option a. values and objectives are intact (more simple), while option b. says that values and objectives are affected by developmental problems. We could tentatively start from option a. (even if it is not possible to ignore the option b.!).

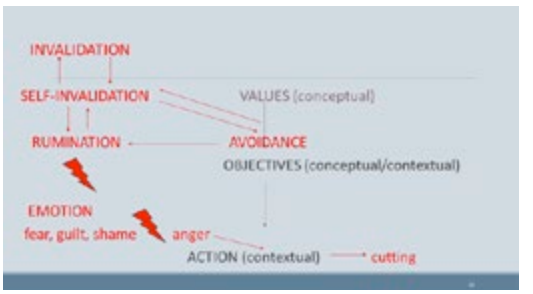
The first step is what happens when an intense emotion, for instance fear, takes the control of behaviour in a certain situation. The result could be a self-damaging action, such as self-cutting. A very high percentage of BPD subjects use to cut themselves, for instance in difficult social situations, mainly when they feel the risk of social exclusion, or abandonment. The result of this kind of action is a momentary emotional relief. This is also demonstrated from a neurobiological perspective: indeed we know that the amygdala, a part of the brain that is activated by emotionally relevant stimuli, is often over-activated in BPD subjects and that physical pain has the capability to reduce its over-activation.

However relief is short because secondary emotions suddenly happen. Guilt or shame are typically secondary emotions raised by dysfunctional behaviour. A typical pattern happens in emotional eating, where binge eating is often caused by intense anxiety. Eating a lot of food helps to reduce the emotion, however both physical and psychological discomfort contribute to the appearance of secondary emotions. The problem here is what to do with secondary emotions.

A possible solution is behavioural avoidance, related to both values and objectives. Avoidance has recently been considered as an attempt to downregulate emotions in BPD. However

avoidant behaviour has serious consequences, mainly if we now take into consideration the option b. that we mentioned before. Option b. says that to become BPD it is necessary a long, problematic, developmental history. Linehan's model, previously mentioned, assumes that emotional vulnerability and environmental invalidation are two components of a negative developmental path. However it seems that to tolerate emotional invalidation future BPD subjects use to self-invalidate themselves. Self-invalidation is an attempt to cope with two poles of a dilemma: emotional dysregulation and environmental invalidation. Since the last one increases the former, and vice-versa, self-invalidation is a way to be accepted by an invalidating environment. To be concrete: if somebody tells me that I'm wrong because he/she doesn't understand my emotional difficulties, I feel invalidated and my negative emotions are at risk to increase. If I convince myself that I'm wrong, the invalidating environment is satisfied, however I don't regulate my emotions. Self-invalidation makes me weaker: I start to consider myself not adapted to life. The risk of avoidance is to increase self-invalidation and to reduce self-efficacy.
Cognitive rumination is a further attempt to cope with negative emotions raised by avoidance, self-invalidation, environmental invalidation. However rumination produces emotions such guilt and shame, that could finally change into anger. It is well known that to be angry is "better" than to be guilty, ashamed or anxious. From a subjective perspective to be angry means to be stronger. The final result is, unfortunately, that anger is finally self-directed. That is that it determines the possibility to act in a self-aggressive way. The vicious circle is so completed.
The figures summarise what is explained in the text.





REFERENCES

Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010), "Emotion regulation strategies across psychopathology: A meta-analytic review", *Clinical Psychology Review,* 30, pp. 217-237;
APA (2000). *Diagnostic and statistical manual of mental disorders*, *4*;
Bargh, J. A., and Williams, L. E. (2007) "The nonconscious regulation of emotion", In J.J. Gross (ed.), *Handbook of emotion regulation,* Guilford Press, New York, pp. 429-445;
Beblo, T., Fernando, S., Kamper, P., Griepenstroh, J., Aschenbrenner, S., Pastuszak, A., Schlosser, N., and Driessen, M. (2013), "Increased attempts to suppress negative and positive emotions in Borderline Personality Disorder", *Psychiatry research*, 210(2), pp. 505-509;
Carpenter, R. W., and Trull, T. J. (2013), "Components of emotion dysregulation in borderline personality disorder: a review", *Current psychiatry reports*, 15(1), pp. 1-8;
Chapman, A. L., Dixon-Gordon, K. L., and Walters, K. N. (2011), "Experiential avoidance and emotion regulation in borderline personality disorder", *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 29(1), pp. 35-52;

Dixon-Gordon, K. L., Chapman, A. L., Weiss, N. H., and Rosenthal, M. Z. (2014), "A preliminary Examination of the role of emotion differentiation in the relationship between borderline personality and urges for maladaptive behaviors", *Journal of psychopathology and behavioral assessment*, 36(4), pp. 616-625;
Doll, A., Sorg, C., Manoliu, A., Woller, A., Meng, C., and Forstl, H. (2013), "Shifted intrinsic connectivity of central executive and salience network in borderline personality disorder", *Front Hum Neurosci*, 7, p. 727;
Donegan, N. H., Sanislow, C. A., Blumberg, H. P., Fulbright, R. K., Lacadie, C., Skudlarski, P., Gore, J. C., Olson, I. R., McGlashan, T. H., and Wexler, B. E. (2003), "Amygdala hyperreactivity in borderline personality disorder: implications for emotional dysregulation", *Biological psychiatry*, 54(11), pp. 1284-1293;
Fairholme, C.P., Boisseau, C.L., Ellard, K.K., Ehrenreich, J.T., and Barlow, D.H. (2010), "Emotions, emotion regulation, and psychological treatment: A unified perspective", In Kring, A. M., and Sloan, D. M. (eds.), *Emotion regulation and Psychopatology: A transdiagnostic approach to etiology and treatment,* Guilford Press, New York, pp. 283-309;
Gill, D., and Warburton, W. (2014), "An investigation of the biosocial model of borderline personality disorder", *Journal of clinical psychology*, 70(9), pp. 866-873;
Glenn, C. R., and Klonsky, E. D. (2009), "Emotion dysregulation as a core feature of borderline personality disorder", *Journal of personality disorders*, 23(1), pp. 20-28;
Górska, D. (2013), "Emotions (un) expressed in words: referential activity in borderline personality disorder", *Acta Neuropsychologica*, 11(2), pp. 143-160;
Gross, J. J., Richards, J. M., and Jhon, O. P. (2006), "Emotion regulation in everyday life", In Snyder, D. K., Simpson, J. A., and Hughes, J. N. (eds.), *Emotion regulation in couples and families: Pathways to dysfunction and health,* American Psychological Association, Washington DC, pp. 13-35;
Gross, J. J., and Thompson, R. A. (2007), "Emotion regulation: conceptual foundations", In Gross, J. J. (ed.), *Handbook of emotion regulation,* Guilford Press, New York, pp 3-24;
Kuo, J. R., and Linehan, M. M. (2009), "Disentangling emotion processes in borderline personality disorder: physiological and self-reported assessment of biological vulnerability, baseline intensity, and reactivity to emotionally evocative stimuli", *Journal of abnormal psychology*, 118(3), p. 531;
Kuo, J. R., Neacsiu, A. D., Fitzpatrick, S., and MacDonald, D. E. (2014), "A methodological examination of emotion inductions in borderline personality disorder: a comparison of standardized versus idiographic stimuli", *Journal of Psychopathology and Behavioral Assessment*, 36(1), pp. 155-164;
Linehan, M. (1993), *Cognitive-behavioral treatment of borderline personality disorder*, Guilford Press, New York;
Loewenstein, G. (2007), "Affect regulation and affective forecasting", In Gross, J. J. (ed.), *Handbook of emotion regulation,* Guilford Press, New York, pp. 180-203;
Mennin, D. S., and Fresco, D. M. (2009), "Emotion regulation as an integrative framework for understanding and treating psychopathology", *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment*, pp. 356-379;
Newhill, C. E., Eack, S. M., and Mulvey, E. P. (2012), "A growth curve analysis of emotion dysregulation as a mediator for violence in individuals with and without borderline personality disorder", *Journal of personality disorders*, 26(3), p. 452;
Pietrek, C., Popov, T., Steffen, A., Miller, G. A., and Rockstroh, B. (2012), "Neuromagnetic indication of dysfunctional emotion regulation in affective disorders", *Depression research and treatment*, *2012*;
Putnam, K. M., and Silk, K. R. (2005), "Emotion dysregulation and the development of borderline personality disorder", *Development and psychopathology*, 17(04), pp. 899-925;

Reeves, M., James, L. M., Pizzarello, S. M., and Taylor, J. E. (2010), "Support for Linehan's biosocial theory from a nonclinical sample", *Journal of personality disorders*, 24(3), p. 312;
Ruocco, A. C., Amirthavasagam, S., and Zakzanis, K. K. (2012), "Amygdala and hippocampal volume reductions as candidate endophenotypes for borderline personality disorder: a meta-analysis of magnetic resonance imaging studies", *Psychiatry Research: Neuroimaging*, 201(3), pp. 245-252;
Sansone, R. A., and Sansone, L. A. (2010), "Emotional hyper-reactivity in borderline personality disorder", *Psychiatry (Edgmont)*, 7(9), p. 16;
Schmahl, C., Herpertz, S. C., Bertsch, K., Ende, G., Flor, H., Kirsch, P., Lis, S., Meyer-Lindenberg, A., Rietschel, M., Schneider, M., Spanagel, R., Treede R.-D., and Bohus, M. (2014), "Mechanisms of disturbed emotion processing and social interaction in borderline personality disorder: state of knowledge and research agenda of the German Clinical Research Unit", *Borderline personality disorder and emotion dysregulation*, 1(1), p. 1;
Selby, E. A., and Joiner Jr, T. E. (2009), "Cascades of emotion: The emergence of borderline personality disorder from emotional and behavioral dysregulation", *Review of General Psychology*, 13(3), p. 219;
Vine, V., and Aldao, A. (2014), "Impaired emotional clarity and psychopathology: A transdiagnostic deficit with symptom-specific pathways through emotion regulation", *Journal of Social and Clinical Psychology*, 33(4), p. 319;
Zaki, L. F., Coifman, K. G., Rafaeli, E., Berenson, K. R., and Downey, G. (2013), "Emotion differentiation as a protective factor against nonsuicidal self-injury in borderline personality disorder", *Behavior therapy*, 44(3), pp. 529-540.

FLAVIA FELLETTI
University of Barcelona
Flavia.felletti@gmail.com

WHAT AUTISM CAN TELL US ABOUT THE LINK BETWEEN EMPATHY AND MORAL REASONING?[1]

abstract

I will discuss the relationship between empathy and moral reasoning among people with autism. I will discuss the deficit that people with autism show in empathy, that affects mostly perspective taking, and the studies conducted by Blair (1996) and Moran et al. (2011), which suggest that people with autism are not significantly impaired in moral reasoning. I will argue that perspective taking might play an important role in moral reasoning. As Moran et al. found, unlike typically developed, people with autism do not judge accidental and attempted harms differently. I will suggest that their deficit in perspective taking might explain this difference. However, I will conclude, studies on autism do not help to assess the influence of the affective components of empathy on moral reasoning.

keywords

empathy, moral reasoning, autism, perspective taking

Phenomenology and Mind, n. 11 - 2016, pp. 222-230
DOI: 10.13128/Phe_Mi-20121
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

One of the hottest topics of discussion in philosophy of mind, psychology, neuroscience, and related disciplines is the link between empathy and morality. On the one hand, authors such as Hoffmann (2000) assign a key role to empathy in morality, appealing mostly to the role that empathy plays in motivating prosocial behavior. On the other hand, Prinz (2011) argues that empathy is not necessary for any aspect of morality, and in particular, that it is not necessary neither for moral development, nor for moral judgment, nor for motivating moral conduct. However, although Prinz might be right in claiming that empathy is not *necessary* for morality, certainly there is a sense in which empathy *does* influence morality. Thus, the real interesting question is *how and to what extent empathy influences morality*, and the answer to this question partly depends on how empathy is conceived.

In Prinz (2011: 212)'s usage of the term, "empathy is a kind of vicarious emotion: it is feeling what one takes another person to be feeling. And the 'taking' here can a matter of automatic contagion or the result of a complicate exercise of the imagination." By contrast, a number of empirical studies are designed in line with Davis's (1980, 1994) characterization of empathy, according to which empathy is a multidimensional construct with cognitive and affective components that are to be measured separately. Following Davis, the affective components of empathy are *empathic concern* (EC), which refers to the dispositional tendency to experience feelings of sympathy, concern, and compassion for unfortunate others, and *personal distress* (PD), which refers to the feeling of personal discomfort, uneasiness, and distress when exposed to the distress of others. While the cognitive component of empathy is *perspective taking* (PT), which refers to the dispositional tendency to entertain the psychological point of view of others.

The distinction between affective and cognitive components of empathy becomes particularly relevant for those studies in which a deficit in empathy and its effects on human behavior are concerned, for it permits to characterize with greater accuracy the nature of the deficit. Much can be learned about the relationship between empathy and morality by focusing

1 A previous version of this paper was presented in the conference *Emotions, Normativity, and Social Life*, organized by the University of San Raffaele. I want to thank the audience of the conference and, in particular, Sarah Songhorian, Cecilea Mun, Bianca Cepollaro, Max Hayward, and Roberta De Monticelli for their useful comments. In addition, I want to thank Carlos Muñoz-Suárez for helpful comments and suggestions on some of the ideas discussed in this paper and the anonymous reviewers of the journal *Phenomenology and Mind* for their comments and suggestions on the first version of the paper.

the attention on those people who show a deficit in empathy. Indeed, studies conducted on people with autism – often characterized as an empathy disorder – highlighted important aspects of the relationship between empathy and moral reasoning. In this paper, I will discuss two studies respectively conducted by Blair (1996) and Moran *et al.* (2011) to investigate moral reasoning among people with autism, and the conclusions that might be drawn about the relationship between empathy and moral reasoning upon the consideration of their findings.

I will proceed as follows. Firstly, I will discuss the issue of empathy among people with autism and the nature of the deficit that people with autism show in empathy, which, as many studies suggest, mostly affects the cognitive component of empathy. I will proceed, then, by discussing the studies conducted by Blair (1996) and Moran *et al.* (2011) and their respective findings. Both studies suggest that, despite their deficit in the cognitive component of empathy, people with autism are not significantly impaired in moral reasoning. However, as the findings of Moran *et al.* show, in judging the moral status of an action people with autism, unlike normally developed, appear to consider the outcome of the action more than the intention of the agent, consequently not judging accidental and attempted harms differently. I will suggest that the deficit that people with autism show in perspective taking might be the key to explain this difference, thus supporting the hypothesis that one component of empathy, perspective taking, influences moral reasoning in an important way. However, I will conclude, since the deficit that people with autism show in empathy appears to affect mostly the cognitive component of empathy, studies on this population might not be helpful to assess the influence of the affective component of empathy on moral reasoning.

1. Addressing the issue of empathy among people with autism

Autism is a pervasive developmental disorder characterized by impairments in social behavior, a varying degree of impairments in communication, and unusually strong, narrow interests and repetitive behaviors (Baron-Cohen 2009). In the literature of philosophy and psychology, autism has often been characterized as an empathy disorder (e.g., Gillberg 1992, Kennett 2002, and Baron-Cohen 2009), and there are at least three arguments used in support of this characterization.

1.1. Autism as an empathy disorder

First, the argument from imitation. Imitation emerges very early in infancy and plays a role in the development of communication and social behavior (Ingersoll 2008). Decety & Meltzoff (2011) proposed that it also plays a role in the development of empathy, as studies in developmental psychology and cognitive neuroscience seem to suggest. Studies in developmental psychology found that infants are able to imitate basic gestures (*e.g.*, tongue protrusion) immediately after birth, which suggests that infants have an innate ability for recognizing an equivalence between the acts of the others and the acts of the self. In line with this, Meltzoff (2007) proposed the 'Like-me' developmental framework of early inter-subjectivity, according to which imitation and the mechanisms underlying it constitute a bridge that allows the infant to connect with the others, understanding their mental states through the observation of their behavior. Studies in cognitive neuroscience (*e.g.*, Carr *et al.* 2003, and Dapretto *et al.* 2006), instead, found that the imitation and the observation of emotional facial expressions activate in great part similar brain areas[2], and that greater activity is registered when the subjects are not only observing but also imitating the facial expressions. Several studies found a deficit in imitation in people with autism. Among these, Sigman & Ungerer (1984) found vocal and gestural imitation impairments in children with autism. While Hobson & Lee (1999) found that although children with autism performed

2 *E.g.*, inferior frontal cortex, superior temporal cortex, insula, and amygdala.

well at imitating the actions, they did not imitate the style (gentle or harsh) of the target, suggesting that the social and affective form of imitation, whose mechanisms might overlap with the ones underlying empathy, is impaired in children with autism (Iacoboni 2009).
Second, and related, the argument from mirror neurons. The mirror neuron system (MNS) is thought to be involved in imitation and in social cognition, and it might mediate our understanding of other people's emotions enabling the translation of an emotional state observed into one of the self (Carr *et al.* 2003, Dapretto *et al.* 2006, Iacoboni 2009). A number of studies found abnormal activation of the MNS system in people with autism[3]. Among these, Dapretto (Dapretto *et al.* 2006) conduced an fMRI study on high-functioning autism and typically developing children while imitating and observing emotional facial expressions. What Dapretto found is that, despite performing well in the imitation tasks, children with autism showed an abnormal activation in the pars opercularis, whose activation has been reported during imitation, action observation, and understanding of intentions. However, as Magnee *et al.* (2007) and South & Hamilton (2008) argue, it is controversial both that the MNS is impaired in people with autism, and that the MNS plays a role in imitation and in empathy. More specifically, they argue that the primary function of the MNS is action prediction, which is not impaired in people with autism, and that these studies might have not considered reduced attention to social stimuli and differences in a general understanding of complex visual information, both reported in people with autism.
Finally, the argument from the Theory of Mind (ToM). Perspective taking relies on ToM related abilities, and people with autism show poor ToM related abilities, as it is supported for instance by the fact that, unlike typically developing, children with autism over the age of 4 normally fail to provide the correct answer in the 'Sally-Anne' false-belief task (Perner & Wimmer 1983, Baron-Cohen *et al.* 1985). In this task, the children are presented with a story involving two characters, Sally and Anne, impersonated by puppets or real people. Sally and Anne are initially in the same room. Sally puts a ball in a basket and covers it, and then she leaves the room. While Sally is outside, Anne moves the ball from the basket into a nearby box. At this point, children are asked 'Where will Sally look for the ball when she comes back to the room?' The children who say that she will look in the box fail to see the situation from Sally's perspective.

1.2. The nature of the deficit

Many studies addressing the issue of empathy among people with autism suggest that the deficit that autistic individuals show in empathy mostly affects the cognitive component of empathy, *i.e.*, perspective taking. Studies in which Davis's Interpersonal Reactivity Index – a questionnaire who measures separately all components of empathy – was employed reported a higher score in the PD subscale, no significant differences in the EC, and a significantly lower score in the PT subscale in people with autism compared to controls (Rogers *et al.* 2007, Dziobek *et al.* 2008, and Hirvela & Helkama 2011). While in those studies designed to measure the psychophysiological responsiveness of people with autism to emotional stimuli normal or enhanced responsiveness was found in people with autism compared to controls. Among these, Blair (1999) recorded the skin conductance activity of people with autism and controls in response to picture of distressed targets and found no significant differences between the two groups. While Magnee *et al.* (2007) recorded the facial electromyographic responses of people with autism and controls following the presentation of visual emotion stimuli (facial expressions) and audiovisual emotional pairs (faces and voices), and found that people with autism show heightened responses to happy and fearful faces, and intact responses to the audiovisual stimuli compared to controls.

3 *E.g.*, Oberman *et al.* (2005) and Bernier *et al.* (2007).

In line with these results, Smith (2009) proposed what he called 'The Empathy Imbalance Hypothesis of Autism' (EIH), according to which people with autism have an enhanced affective empathy and a deficit concerning the cognitive component of empathy. This hypothesis explains, according to Smith, characteristic behavioral traits of autism, such as avoidance, tendency to become over stimulated, and difficulties in making eye contact. Indeed, in accordance with the EIH, people with autism would be excessively sensitive to the emotions of others and vulnerable to empathic overarousal. Empathic oversarousal is defined by Hoffmann (2000: 198) as "an involuntary process that occurs when an observer's empathic distress becomes so painful and intolerable that it is transformed into an intense feeling of personal distress, which may move the person out of the empathic mode entirely." In addition, the EIH would also be in line, as Smith notes, with Markram *et al.* (2007)'s "intense world" hypothesis of autism, according to which a hyper-responsiveness of the amygdala – reported in people with autism, *e.g.*, by Dalton *et al.* (2005) – would make people with autism excessively responsive to socio-emotional stimuli.

2. Autism, Empathy, and Moral Reasoning

Studies investigating moral reasoning among people with autism lead to some important results. In this section, I will discuss two studies, respectively conducted by Blair (1996) and by Moran *et al.* (2011).

Blair (1996) investigated the capacity of children with autism to distinguish between moral and conventional transgressions, a distinction that is typically made by the age of 39 months. Moral transgressions (MTs) and conventional transgressions (CTs) are distinguished on the basis of the following criteria. MTs are defined by their consequences for the welfare of others, while CTs are defined by their consequences for the social order. MTs are regarded as more serious and do not depend on the rule condition (*i.e.*, if the rules change, they are still regarded as impermissible), while CTs are not regarded as transgressions if no rule is violated. Finally, MTs involve a victim, whereas CTs do not. The participants in the experiment were a group of controls and two groups of children with autism: one who passed two simple false-belief tasks (ToM group), thus showing an acquired ability to mentalize, and the other group who failed in both the false-belief tasks (no-ToM group). The ability to distinguish between MTs and CTs was tested by presenting the subjects with four MTs and four CTs stories told by using Playmobil characters. Three kinds of questions were asked in order to assess the subject's ability to evaluate, in turn, the permissibility of the action, the seriousness of the transgression, and the authority's jurisdiction[4]. The findings of Blair's study showed that, unlike psychopaths – who appear to show a deficit in the affective but not in the cognitive components of empathy (Blair 1995, 1997) – people with autism clearly distinguish between MTs and CTs. This might suggest that the affective components of empathy might rather be involved in the making of this distinction, although this issue should be investigated separately. Furthermore, the ability in the false-belief tasks was not associated with the ability to make the MTs/CTs distinction, as it is supported by the fact that also the no-Tom group made it. In line with this, Blair suggested that people with autism are sensitive to the distress of others (in this case, of the victim involved in the MTs) and that the ability to mentalize is not a pre-requisite for making the moral/conventional transgressions distinction.

Moran *et al.* (2011) investigated, instead, whether adults with autism would made atypical moral judgments when they needed to take into consideration the intention of an agent (which requires ToM related abilities) and the outcome of her action (which does not require ToM related abilities). The participants, a group of adults with autism and one of typically

4 Questions were, e.g., 'would it be OK for [the character] to do [the action] if [the authority] said she can?'

developed, were subjected to a simple false-belief task and to a moral judgment task. In the moral judgment task, subjects were presented with vignettes in which a character, either intentionally or accidentally, caused a harm (killed someone) or a neutral outcome, and were asked to evaluate the moral status of the character's action. Four possible situations were illustrated in the vignettes: (i) neutral action (neutral intention, neutral outcome), (ii) attempted harm (negative intention, neutral outcome), (iii) accidental harm (neutral intention, negative outcome), and (iv) intended harm (negative intention, negative outcome). The agent's negative intention was based on the belief that her action was going to kill someone, and *vice versa* the neutral intention was based on the belief that her action was not going to cause any harm. The results showed no difference in the false-belief task performance between participants with autism and controls, thus indicating that the participants with autism had acquired a basic ability to mentalize. Concerning the moral judgment task, a selective difference was found between participants with autism and controls in the judgment of accidental harms. Indeed, unlike typically developed, participants with autism appear to do not judge accidental harms as less morally wrong than attempted harms. Furthermore, in making their judgments, participants with autism appeared to take into account the outcome of the action more than the agents' beliefs and intention. This indicates, according to Moran and colleagues, that people with autism show an impairment in integrating mental state information when making moral judgments.

However, what the findings of Moran *et al.* show is that participants with autism, despite having acquired a basic ability to mentalize, in making moral judgments rely significantly less on information about the agent's mental states. But it is important to notice here that a considerable amount of mental states information must be taken into account; namely, the agent's beliefs and the agent's intention based on such beliefs. Thus, more complex mentalizing abilities seem to be involved in this kind of moral judgments, and such abilities might be the ones involved in perspective taking.

Indeed, the deficit that people with autism show in perspective taking might be the key to explain the difference between participants with autism and typically developed participants. This, because typically developed arguably tend to assume the agent's perspective while judging the moral status of her action, and in particular when judging accidental actions. Because of this, the agent's beliefs and intention play a relevant role in their moral judgments. By contrast, people with autism plausibly fail to take the agent's perspective. Thus, the agent's beliefs and intention play a minor role in their moral judgments and consequently the outcome of her action assumes a significantly greater relevance in their moral judgments. If this is true, then the findings of Moran and colleagues would support the hypothesis that one component of empathy, perspective taking, plays an important role in our moral reasoning. This also seems to be in line with the findings of Young & Saxe (2008a, b)'s fMRI studies, which indicate that brain areas that are typically activated in ToM related tasks are also activated during moral reasoning and suggest that a spontaneous process of mental state inference takes place during moral judgment tasks. However, further studies are needed in order to verify the correctness of this hypothesis.

3. Conclusions

In this paper, I discussed empathy and moral reasoning among people with autism, and some aspects of the relationship between empathy and moral reasoning that studies on people with autism might help us to highlight. The following conclusions appear to emerge from the discussion. Firstly, as many studies suggest, the deficit that people with autism show in empathy mostly affects the cognitive component of empathy, *i.e.*, perspective taking. Because of this, studies on people with autism might not be adequate to assess the influence of the affective components of empathy on moral reasoning. Studies on psychopaths, who appear

instead to show a deficit specifically in the affective components of empathy (Blair 1995, 1997), might rather be helpful to investigate this matter. Secondly, as Blair's (1996) findings show, the ability to mentalize – on which perspective taking relies and that is typically impaired in people with autism – does not represent a pre-requisite for distinguishing between moral and conventional transgressions. Indeed, also the participants who failed the two simple false belief tasks appeared to make this distinction. This result is in stark contrast with findings on psychopaths (Blair 1995, 1997), which suggest that affective components of empathy might rather be necessary for the making of this distinction; however, this matter needs to be investigated separately. Moreover, the findings of Moran *et al.* (2011) also seem to suggest that people with autism are not significantly impaired in moral reasoning. This, because, in the moral judgment task employed by Moran *et al*., the only relevant difference between typically developed and participants with autism concerned the judgment of accidental harms. However, as the findings of Moran and colleagues show, unlike typically developed, participants with autism do not judge differently accidental harms and attempted harms, suggesting that, despite having acquired a basic ability to mentalize, they show an impairment in integrating mental state information for moral judgments. I suggested that the deficit that people with autism show in perspective taking might provide an explanation for this difference. Indeed, more than basic mentalizing abilities might be required in the moral judgment task presented by Moran and colleagues. The participants needed, in fact, to consider both the agents' beliefs and the agents' intentions based on such beliefs. I proposed that the required mentalizing abilities might be the ones involved in perspective taking, which could explain the selective difference found between participants with autism and controls in the moral judgment task as follows. Arguably, in making moral judgments that require considering the agent's intentions, typically developed tend to assume the agent's perspective, thus assigning a more important role to the agent's mental state compared to her action's outcome. By contrast, participants with autism plausibly fail to take the agent's perspective. Thus, in making these kind of moral judgments, they assign a minor role to the agent's mental states and a more relevant role to the outcome of her action, and consequently they do not judge accidental and attempted harms differently. This would support the hypothesis that perspective taking plays an important role in our moral judgments, and in particular in those judgments in which mental states must to be taken into account. However, further studies are needed in order to verify the correctness of this hypothesis.

REFERENCES

Baron-Cohen, S. (2009), "Autism: The Empathizing–Systemizing (E-S) Theory", *The Year in Cognitive Neuroscience 2009,* Ann. N.Y. Acad. Sci. 1156, pp. 68–80;
Baron-Cohen, S., Leslie, A.M., and Frith, U. (1985), "Does the autistic child have a "theory of mind"?", *Cognition*, 21, pp. 37–46;
Blair, R. (1995), "A cognitive developmental approach to morality: investigating the psychopath", *Cognition* 57, pp. 1-29;
– (1996), "Brief Report: Morality in the Autistic Child", *Journal of Autism and Developmental Disorders*, Vol. 26, No. 5, pp. 571-579;
– (1997), "Moral reasoning and the child with psychopathic tendencies", *Personality and Individual Differences,* Vol. 22, No. 5, pp. 731-739;
– (1999), "Psychophysiological responsiveness to the distress of others in children with autism", *Personality and Individual Differences*, Vol. 26, No. 3, pp. 477–485;
Bernier, R., Dawson, G., Webb, S., and Murias, M. (2007), "EEG Mu Rhythm and Imitation Impairments in Individuals with Autism Spectrum Disorder", *Brain Cognition,* Vol. 64, No. 3, pp. 228–237;

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003), "Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas", *Proceedings of the National Academy of Sciences of the United States of America*, 100, pp. 5497–502;
Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., and Davidson, R. J. (2005), "Gaze Fixation and the Neural Circuitry of Face Processing in Autism", *Nature Science*, 8, pp. 519-526;
Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., and Iacoboni, M. (2006), "Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorder", *Nature Neuroscience*, Vol. 9, No.1, pp. 28-30;
Davis, M. H. (1980), "A Multidimensional Approach to Individual Differences in Empathy", JSAS *Catalog of Selected Documents in Psychology*, 1980, 10, p. 85;
Davis, M. H. (1994), "The Heritability of Characteristics Associated with Dispositional Empathy", *Journal of Personality*, 62.3, Duke University Press, Durham NC;
Decety, J., and Meltzoff, A. N. (2011), "Empathy, Imitation, and the Social Brain", In Coplan, A., and Goldie, P. (eds.), *Empathy: Philosophical and Psychological Perspectives*, Oxford University Press, Oxford, pp. 58-81;
Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., and Convit, A. (2008), "Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (MET)", *Journal of Autism and Developmental Disorders*, 38, pp. 464-473;
Gillberg, C. L. (1992), "Autism and autistic-like conditions in subclasses among disorders of empathy", *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 33, pp. 813–842';
Hirvelä, S., and Helkama, K. (2011), "Empathy, Values, Morality, and Asperger's Syndrome", *Scandinavian Journal of Psychology*, 2011, 52, pp. 560–572;
Hobson, P., and Lee, A. (1999), "Imitation and Identification in Autism", *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* Vol. 40, No. 4, pp. 649-659;
Hoffman, M. L. (2000), *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge University Press, Cambridge;
Iacoboni, M. (2009), "Imitation, Empathy, and Mirror Neurons", *Annual Review of Psychology*, 60, pp. 653–70;
Ingersoll, B. (2008), "The Social Role of Imitation in Autism: Implications for the Treatment of Imitations Deficits", *Infants&Young Children,* Vol. 21, No. 2, pp. 107-119;
Kennett, J. (2002), "Autism, Empathy, and Moral Agency", *The Philosophical Quarterly*, Vol. 52, No. 208, pp. 340-357;
Magnee, M. J. C. M., De Gelder, B., Van Engeland, H., and Kremner, C. (2007), "Facial electromyographic responses to emotional information from faces and voices in individuals with pervasive developmental disorder", *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 48, pp. 1122–1130;
Markram, H., Rinaldi, T., and Markram, K. (2007), "The intense world syndrome: An alternative hypothesis for autism", *Frontiers in Neuroscience*, *1*, pp. 77–96;
Meltzoff, A. N. (2007), "'Like-Me': a foundation for social cognition", *Developmental Science* 10:1, pp. 126-134;
Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., and Gabrieli, J. D. (2011), "Impaired theory of mind for moral judgment in highfunctioning autism", *Proceedings of the National Academy of Sciences*, Vol. 108, No. 7, pp. 2688–2692;
Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005), "EEG evidence for mirror neuron dysfunction in autism spectrum disorders", *Cognitive Brain Research*, 24, pp. 190–198;

Perner, J., and Wimmer, H. (1983), "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception", *Cognition*, 13, pp. 103-108;
Perner, J., Frith, U., Leslie, A. M., and Leekam, S. R. (1989), "Exploration of the autistic child's theory of mind: Knowledge, belief and communication", *Child Development,* 60, pp. 689-700;
Prinz, J. (2011). "Is Empathy Necessary for Morality?", In Coplan, A., and Goldie, P. (eds.), *Empathy: Philosophical and Psychological Perspectives*, Oxford University Press, Oxford, pp. 211-229;
Rogers, K., Dziobe, I., Hassenstab, J., Wolf, O. T, and Convit, A. (2007), "Who Cares? Revisiting Empathy in Asperger Syndrome", *Journal of Autism and Developmental Disorders*, 37, pp. 709–715;
Sigman, M., and Ungerer, J. A. (1984), "Cognitive and Language Skills in Autistic, Mentally Retarded, and Normal Children", *Developmental Psychology*, Vol. 20, No. 2, pp. 293-302;
Smith, A. (2009), "The Empathy Imbalance Hypothesis of Autism: A Theoretical Approach to Cognitive and Emotional Empathy in Autistic Development", *The Psychological Record*, 59, pp. 489-510;
Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., and Saxe, R. (2010), "Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments", *PNAS*, Vol. 107, No.15, pp. 6753-6758;
Young, L., and Saxe, R. (2008a), "The neural basis of belief encoding and integration in moral judgment", *Neuroimage*, Vol. 40, Issue 4, pp. 1912-1920;
Young, L., and Saxe, R. (2008b), "An fMRI investigation of spontaneous mental state inference for moral judgment", *Journal of Cognitive Neuroscience*, 21, 7, pp. 1396–1405;
Williams, J. H. G., Whiten, A., Suddendorf, T., and Perrett, D. I. (2001), "Imitation, mirror neurons and autism", *Neuroscience and Biobehavioral Reviews*, 25, pp. 287-295.

LAURA CANDIOTTO
University of Edinburgh
Laura.Candiotto@ed.ac.uk

EXTENDED AFFECTIVITY AS THE COGNITION OF PRIMARY INTERSUBJECTIVITY[1]

abstract

I discuss the primordial affectivity approach (Colombetti 2014) and the extended emotions theory (Krueger 2014, Slaby 2014, Candiotto 2015, Carter et al. 2016) in order to propose a novel account of "extended affectivity" (EA) as the cognition of primary intersubjectivity (EACPI). I explain why the distributed cognition model is the more convenient to understand the collective and the subjective dimension of EA. The novelty of EACPI consists in the recognition of the protocognitive valence of the affectivity, referring to the work of Colwin Trevarthen (Trevarthen 1979; Trevarthen 2011), who has demonstrated the leading role of affectivity in the neonatal intersubjectivity in neurobiology.

keywords

affectivity, primary intersubjectivity, distributed cognition, extended emotions, enactivism

Phenomenology and Mind, n. 11 - 2016, pp. 232-241
DOI: 10.13128/Phe_Mi-20122
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

Among the many novel ongoing approaches in the contemporary debate about the nature of emotions, the primordial affectivity approach (PAA) (Colombetti 2014) and the extended emotions theory (EET) (Krueger 2014, Slaby 2014, Candiotto 2015, Carter *et al.* 2016) stand out for their capacity to combine a well-defined theoretical account of the mind with the phenomenology of affectivity and the empirical works coming from cognitive science and neuroscience. Moreover, they envisage extending their scope of applicability to other fields of research, such as social psychology, sociology and technology. PAA and EET come from two different approaches to the mind – notably, active externalism and enactivism – but they have in common the recognition of the constitutive value of the environment for the realization of cognition as affectivity. In this paper, I will discuss these approaches in detail in order to propose a novel account of "extended affectivity" (EA) as the cognition of primary intersubjectivity (EACPI). Affectivity is the most primitive way in which the subject knows the environment in which she/he is embedded, and the first dynamic practice to constitute a "world in common" (Nancy 2002).

1. Contemporary trends in the theory of emotions

It is possible to analyse contemporary trends in the theory of emotions by dividing them into three broad groups which are not, naturally, exhaustive and fine-grained (cf. Candiotto 2016a). The first approach is cognitivism. In the philosophy of emotion, cognitivism assumes the propositional attitude identity thesis. This thesis states that emotions are identical to propositions, or in its weaker form, the thesis maintains that emotions involve propositional attitudes. The specific judgement made by emotions is defined as "appraisal", i.e. a judgement of value. The propositional attitude identity thesis derives from a critique of (1) feeling-centered approaches to emotions, which conceptualise emotions as bodily feelings, (2) and behaviourist programs, which are understood as inadequate to fully explain emotions' intentionality.

In contrast, within the cognitive framework, emotions are understood as occurrent states of mind with a specific intentionality. They are different from moods, which are occurent states

1 This paper arises from the project "Emotions First", funded by the EU (Marie Curie Individual Research Fellowship, grant number: 655143), I'm carrying on at the Eidyn Research Centre of the University of Edinburgh. I would thank those with whom I had the opportunity to discuss my approach, in particular Adam J. Carter, Jessica MacLaren, and Dory Scaltsas for their insightful comments.

but with a general intentionality, and from sentiments and traits as dispositional states. Appraisal is also central to the second approach, the perceptual model. In this model, emotions are types of perceptions, and as perceptions are related to judgment about the empirical word, emotions are therefore related to moral judgments (Goldie 2007, Sauer 2012). As the standard cognitive approach, the perceptual model emphasizes the "feeling towards", i. e. the intentional character of emotions, making a comparison between emotions and perceptions: just like perceptions, emotions overcome themselves in order to reach the object they are for. For this model, emotions "need not consist in articulated propositions" (De Sousa 2014: 19) and, thus, the non conceptual apprehension of the world of beasts and babies (cf. Deigh 1994) is comprehended by binding emotions with desires.

The third group is the most multifaceted and comprises approaches that identify emotions with bodily experience or that stress the strong value of body experience in the emotional arousal. Even if these approaches have some peculiar traits creating particular differences among them, one could still highlight the common rejection of the standard assumption that cognition is instantiated "centrally" by the brain only. Broadly speaking, in this third category we could count those approaches as emanating from Continental philosophy, mainly from phenomenology, existentialism and women's philosophy. Arguably, William James may be seen as one of the founders of this varied family of approaches, since for him emotions are bodily feelings. In the apprehension of reality the bodily feeling comes first, and then the judgment of experience follows. Physiological changes precede emotions that are the subjective experience of body changes (James 1950: 173).

Outside of these three broad groups, there are many other contemporary approaches to emotions theory. Most notable are multidisciplinary approaches that combine different disciplines such as psychology, cognitive science, philosophy, epistemology, anthropology, sociology, biology and neuroscience. These approaches have the aim of providing a broader and more comprehensive model for understanding emotions. Usually, these approaches frame emotional experiences within a broader cognitive environment and can be similar to the second or the third group, or even offer a novel approach to a reformatted cognitivism.

2. Primordial Affectivity and Extended Emotions

The frameworks of the two approaches I discuss in this paper, i.e. enactivism and the extended mind hypothesis, appertain exactly to this novel and diversified multidisciplinary context. Enactivism (Varela, Thompson, and Rosch 1991; Thompson 2007), by underlying the continuity between mind and nature, creates the grounds for embodied and embedded cognition. The extended mind hypothesis (Clark and Chalmers 1998), by focusing on the cognitive valence of external tools, enlightens the mutual cognitive actions inside an extended cognitive system. Both assume a critical stance towards internalism, arguing for an active externalism where cognitive agents constitute themselves in relation to the environment whilst perception/cognition are active forces of the sense making.

Enactivism and the extended mind hypothesis were chosen as frameworks by two very promising and novel hypotheses in affective science and philosophy of emotions discussed in detail below, the Primordial Affectivity Approach (PAA) and the Extended Emotions Thesis (EET).

PAA represents the theoretical model for a very detailed analysis of enactive affectivity delivered by Colombetti in connection with the new results coming from affective science. PAA grounds the comprehension of emotions in an account of the mind that emphasizes its embodied and affective character, understanding affectivity as the primordial way in which an organism understands, decides and acts in a particular environment. Criticizing the standard conception of cognitive science that understands emotions and moods as transitional state of minds (emotions very quick, moods longer), Colombetti comprehends the mind as always

affective in its origin, since affectivity is understood as the structural pattern in which every form of life produces itself and its world. Emotional events are expressions of the whole living organism embedded in the world. Thus, PAA is a dynamic systems approach to affectivity, in which sense-making is the leading notion for understanding the affective structural relationship which all living systems have with the environment. For Colombetti, saying that affectivity is primordial means that it pertains to all living beings – including those without a neural system – because affectivity is enacted by the whole organism. In this way she strongly criticizes the mainstream neuro-centric thinking about mind in general, and emotions in particular.

EET is understood in different ways. Joel Krueger (2014) discusses the hypothesis of the "environmentally extended emotions" for which there are cases in which the emotions extend the body boundaries and are constituted by external factors. As it was well noticed by Achim Stephan, Sven Walter, and Wendy Wilutzky (2014), not all influences performed by the environment might be understood as extension. On the contrary, two very specific traits should be present – at least in the orthodox version of the extended mind hypothesis – the internal representations and the active structuring of the environment. External factors may also be represented by other human beings – i.e. not only in technological tools – and, thus, Jan Slaby (2014) understands "extended emotions" as a case of "collective emotions", or with Krueger's words as "collectively extended emotions", within the so-called "third wave" of the extended cognition. Emotions should be intended both as emotions that are "common" among the members of a group, and as emotions that are constituted by all the members of a group at the same time, developing what Slaby calls "the political philosophy of mind", i.e. the situated and engaged stance in the philosophy of mind not restricted to abstract theoretical investigations.

Following Slaby's conceptualization of EET, Candiotto (2015) stresses how the mind's extension is not only a question of location but also of determining the type of knowledge that is realized: such knowledge is not just "shared" within a group, but it is also extended in the sense of enhanced or maximized by a collective inquiry. In this precise aspect, this formulation of EET found in the Adam Carter, Emma Gordon and Orestis Palermos' approach a common root. Their goal is to understand EET as a novel application of the hypothesis of extended cognition (HEC), i.e. the claim that some cognition extends beyond skin and skull to parts of the external world. Carter *et al* argue for a fruitful combination between HEC and cognitivist approaches to emotion. According to this approach, the cognitions on which emotions supervene are dynamic and extended processes.

3. Primordial affectivity cum and versus extended emotions

The common ground between these two approaches is represented by the general view of an embodied, situated and environmentally active mind.

Colombetti and Slaby subscribe to the idea of an affective intentionality, probably because of their common phenomenological roots (the Merleau-Pontyian lived body) and Slaby's notion of "interactive coupling" ("the continuous interaction with some expressive environmental structure", Slaby 2014: 37) makes his formulation of EET as enactive (his *motto* is "enaction rules extension"), exactly as Colombetti does, merging "primordial affectivity" and "sense making".

Phenomenology and enactivism are not explicitly considered in the extended mind hypothesis[2] and in the more orthodox formulation of extended emotions by Carter *et al*, which

2 Andy Clark (1997) is one of the founder of the model of embodied cognition and he dealt with the embodied mind and enactivism throughout the years (cf. Clark 2016). Therefore, this division should not be understood rigidly.

are framed by novel discoveries in the cognitive science field. Nevertheless, this does not mean that they are incompatible, since scientific discoveries could provide evidence of our phenomenological experiences, and conversely, our phenomenological experiences could stimulate new scientific research bringing them new intuitions. Therefore, Slaby discusses the extended mind hypothesis – he undertakes the "integrationist" second wave approach of Menary (2010) and not the parity principle of Clark and Chalmers (1998) – and one of Colombetti's main goal is to bring the philosophical idea of primordial affectivity to the contemporary debate in affective science, and so to shape primordial affectivity into scientific results. In particular, Colombetti (2013) advances the "neuro-physio-phenomenology" for the scientific study of emotion experience, combining phenomenological first-person data about emotional feelings and third-person data about mind and body activity. Thus, from a methodological point of view, it is possible to underline the common effort of PAA and EET to overcome the division between humanistic and scientific research, as in the case for many contemporary philosophical researches that embrace the idea of a naturalized philosophy.
Both PPA and EET reject psychological internalism or mentalism, and understand appraisal as constituted by the activity of the whole situated organism. For PPA it is not possible to distinguish appraisal from emotion and, following this path, I depict extended emotions as epistemic emotions, enlightening the inextricable bond between emotions and cognition, starting from the neuroscience account which demonstrates how cognitive and emotional functions are so deeply integrated in the brain (Freeman 2000, Lewis 2005). At the same time, it is crucial to understand what "constituted by" means within these two models as it is exactly in this point that they are arguably very different. I will highlight this by analysing some specific contrasts.
First of all I must emphasize that, unlike PAA, EET is not an already well-developed and unified theory. Therefore the discussion below may miss some of the specific notions that pertain to each single approach within EET
On the one side (Colombetti, Slaby, Candiotto) we find the idea that emotions belong to a lived body that acts inside the environment, on the other (Carter *et al.*) the idea that emotions supervene on mental states and that, even if extended in an environment, they pertain mainly to the brain activity. In fact, the circularity between brain and external environment (Carter *et al.*' EET) and enactive coupling are not at all synonymous. In fact, following the standard formulation of extended mind by Clark and Chalmers (1998), the circularity between brain and environment is understood as causality, for which not only the brain depends on the environment but part of the environment constitutes the cognitive activity of the brain. Instead of circular causality, enactive coupling points to the inextricable and interdependent structural relation between an organism and its environment and cognition is exactly understood as the dynamical product of this relation.
According to enactivism, cognition is realized ("enacted") by the whole living organism embedded in the world and so primordial affectivity represents for Colombetti the specific relation that every form of life (also the simplest one, as bacteria) has with its environment. Thus, affectivity could be understood as the self-organizing pattern of a living organism, modelled by the continuous interchange with the environment. Also EET stresses the constitutive relation with the environment as source of the emotional experience, but for AAP the agents are every living being and for EET not, since it describes the emotional cognition of an evoluted human brain and pushes its discoveries to research in artificial intelligence and in sociological research about the relations among humans. This does not mean that AAP is not interested in the study of humans' affectivity but that its account on the more complex emotional experiences of the humans are framed into the basic structure of life and are understood as highly integrated configuration of the same process of self-organisation.

As I have explained above, for AAP affectivity pervades the mind, it is not just an occurrent state of a "neutral" mind. EET, on the contrary, assumes the standard cognitive analysis of emotions as occurrent state of minds and so extended emotions are the very specific and temporally determined states of mind that are realized just in the very determined conditions of the cognitive integration with the environment. This does not mean that EET is less extreme than AAP: in fact, EET claims that emotions overcome the feeling body since extended emotions are beyond the brain *and* the body. I think that AAP could not accept this conclusion since it frames affectivity exactly as structure of a body. Even if in the conclusion of her book (2014, chapter 7) Colombetti has asserted that enactive emotions do not deny relationship with others, she has nevertheless mainly described them from the perspective of an individual organism. She has not argued for interrelatedness or intersubjectivity as primitive (see par. 4 for my thesis about it).

The last point of difference I would underline has to do with the metaphysical assumption undertaken (explicitly or implicitly) by the two approaches. Even if EET claims to be metaphysical neutral (a functionalist approach), I think that it is possible to detect in it a realist approach, broadly speaking, both ontic and epistemic. On the contrary, for PAA it becomes meaningless to speak about "internal world" and "external world" since the action is realized (enacted) in a fully integrated environment understood as the whole. For AAP cognition is a form of interpretation, while in contrast, EET struggles for an objective knowledge framed by the results of science. Therefore, it is notable that the different premises of the two approaches, a functionalist cognitivism even reformed, and continental phenomenology combined with enactivism, have a strong impact in AAP and EET, making them very dissimilar despite having in common the critiques of internalism and the valorization of the structural relation with the environment. In the meantime, we should not assume their incompatibility.

4. Extended affectivity

Arguably, it is fruitful to combine the core idea of PAA, i.e. the understanding of affectivity as the whole experience of a living body, with the EET claim about extended cognition, in order to achieve the novel notion of "extended affectivity" (EA). With EA, extended cognition is not only the extension of mental states but the cognition of the entire living body, where the whole of an individual's experience is framed within intersubjectivity, understood as primitive. EA poses a lot of challenges to PAA and EET, in particular to PAA to dismiss the notion of complete autonomy of the living organism, while EET is challenged to be open to wider horizons of inquiry connected to the idea of the extended consciousness (Noë 2004, 2010) and to a more liberal interpretation – as it is defined by its proponent – of the extended mind hypothesis, i.e. the "socially extended mind" (Gallagher 2013), or third wave. At the same time EA recognizes its debit to both PAA and EET, in particular the constitutional relationship with the environment benefits from the results of PAA, as with the notion of cognitive extension from EET. Thus, my constructive *motto* is to combine the benefits whilst reducing the limits of the two, with the purpose to propose a new approach that will not merely be a sum of the two, but it would even recognize the strong dependency between them.

To clarify: the strength I would take from EET is the core idea of extension and from PAA the critique to reduce affectivity to an "emotional occurrent event". The area I would improve is the social and intersubjective dimension of the extension, focusing not only on the extension in technological tools but developing the idea of primordial affectivity within intersubjectivity as primitive.

EA is not a totally novel conception since it has to do with the philosophical ideas coming from phenomenology and existentialism that the experience is always shared and mutual and that agency is understandable as "we-intentionality", and from social ontology and epistemology

where the "we" is the main indicator of social action, which implies a "plural subject" and not a simple sum of two or more individuals (Gilbert 1989). In the meantime, the word "extended affectivity" starts to be detectable in few very new articles (Colombetti 2015, Fuchs & Koch 2014) on philosophy of emotions. Nevertheless, the novelty of my account consists in the recognition of the protocognitive valence of the affectivity through the conceptualization of social developmental and psychological researches. I simultaneously point to the necessity of describing "affectivity" as this specific protocognitive ability, overcoming the dualism between cognitive and non-cognitive mental states, and also to recognize intersubjectivity as primitive.

4.1. Primary intersubjectivity

I assume the notion of EA as the cognition of primary intersubjectivity (EACPI). Much has been done in primary intersubjectivity, especially by social and developmental psychology. Specifically, using the notion of "primary intersubjectivity", I am referring to the pioneer work of Colwin Trevarthen (Trevarthen 1979, Trevarthen 2011), who has demonstrated the leading role of affectivity in the neonatal intersubjectivity (but also in the belly!) in neurobiology.

> [...] they show themselves capable of making an effort of will and attentiveness to take part in an emotionally charged reciprocation of arbitrary ways of moving, and so to become part of a dramatic narration of being in companionship with another person. (Trevarthen 2011: 124)

Joel Krueger (2013) has already underlined how the physical interventions caregivers use to regulate infant attention and emotion are part of the infant's socially extended mind; in my opinion the best goals that Krueger's paper has attained were to have stressed how the cultural norms are mediated by the body actions and, in the meantime, to have underlined the key role performed by the body in the cognitive development.
The best example of EACPI is the mutual relationship between a mother and her infant (see par. 4.2). Nevertheless, following Shaun Gallagher (2001)[3], I do not think that EACPI pertains only to mothers and infants but that continues to be active in all social animals at all life stages, as the primordial source of intersubjectivity.
Krueger (2014) has underlined that collective extended emotions are easily detected in infancy, but hardly in adults – even if he provides some examples from music and dance – for which the notion of "individual extended emotions" is more useful, i.e. the recognition of the constitutive role of external factors for the individual emotional arousal. On this point my approach to EA would underline that, even if adults' "individual extended emotions" are more prominent, this does not mean that the collective stance may be not at work on the back as primordial source of intersubjectivity. With the development of the cognitive skills, primary intersubjectivity will be combined with the secondary and the tertiary intersubjectivity but its role cannot be denied in the adults intersubjectivity, especially regarding the researches on empathy. Anya Doly (2014) claims that primary intersubjectivity is the empathic responsiveness to the others[4]. Following this path, I assert that primary intersubjectivity plays in the background in adults, and has an extraordinary role in our mutual desire to constitute a "world in common" (Nancy 2002).
The biological constitution of intersubjectivity, in the meantime, should be combined with

3 "Primary, embodied intersubjectivity is not primary simply in developmental terms. Rather, it remains primary across all face-to-face intersubjective experiences,[...]". Gallagher 2001: 91.
4 "Primary empathy is affective reversibility, which renders us susceptible to another's gaze, to their voice, to their touch." Daly 2014: 237.

the recognition of the first-person experience. Regarding HEC, Orestis Palermos and Duncan Pritchard (Palermos & Pritchard 2013) designed a model that seeks to avoid of the extremism of conceptualization of knowledge as realized only "inside" or "outside" the boundaries of the skin. To achieve it, they combine research on virtue epistemology with the distributive cognition framework for group knowledge, valorizing the specific cognitive skills of every agent in the distributed process. Even if they do not explicitly take emotions into account in their model, I do think that it is a very useful way of understanding the peculiar interweaving between collective and first-person affective experience. In my opinion the distributive cognition hypothesis might strengthen the intersubjectivity vision of emotions for which the emotional embodied experience of a subject is not only affected by but also strengthened by the relationships with others. Emotions are not private state of minds but active and dynamic processes between subjects. They are the forthcoming of our active interaction with the world, clarifying how the regulation of the self always involves the others, and without losing the ontological status of the subject. Understanding EA for adults within this framework permits to achieve the notion of "distributed affectivity" (DA), i.e. this specific affective state of the group, and of its single components. Arguably, DA may develop the idea of collectively extended emotions without neglecting their first-personal character.

4.2. The constitutive relationship between the mother and her fetus

Let analyze in details our key case of EACPI. We should notice how the mother and the fetus[5], and then the infant, are bound together by a constitutive relation moved by the biological needs of survival and development. They form a group ("common intercorporality", Fuchs & De Jaegher 2009) understandable as a single entity or as a systemic dyad in its developmental goals. For the infant this relationship is constitutive of her/his growth in terms of evolutions, adaptations and survival in a specified environment. The affective relation between mother and infant, established in the repeated actions of care performed by the mother, builds the protocognition of the infant. As Trevarthen has demonstrated, infants are born with a natural propensity to intersubjectivity,

> [...] with motives and emotions for actions that sustain human intersubjectivity. [...]. Their Intelligence is prepared to grow and be educated by sharing the meaning of intentions and feelings with other humans by means of many expressive forms of body movement that may be perceived in several modalities. (Trevarthen 2011: 121).

They are constitutively bound together not only for biological reasons – think of the expression "you are the blood of my veins, you are bone of my bone" – but also for social reasons. Intelligence has evolved in social animals for actions to be shared socially[6]. It is well known that the symbiotic relations between a mother and infant could also become pathological for the mother (cf. for example the depression post partum) and for the development of the infant's autonomy; nevertheless, it is in its good performance that the infant learns to be a social being.
Arguably, the constitutive "use" of the mother by the infant as a source of care, nutrition, etc. makes their relationship a case of extended affectivity as extended protocognition. The constitutive relation with the mother constitutes the affectivity of the infant and some

5 I'm referring here to the natural mother and not to unspecified caregivers to stress the biological link at work. Nevertheless, doing so I do not deny the possibility to detect these characteristics also in relations to other caregivers, in particular understanding the crucial role performed by the social practices towards our biological patterns.
6 "The creative output of natural intelligence in moving animal bodies, its conation and emotion", Trevarthen 2011: 122.

protocognitive states arise exactly in this process on "interbodily resonance" (Fuchs & Koach 2014). The love and care of the mother for the infant bounces back to the infant producing not only attachment but also a feeling of affection for her/his source of survival and, thus, a protocognition of intersubjectivity as relational knowing. Claiming that what is extended is affectivity means that this extension is not occasional and momentary[7] (as would be if what is extended was just emotion as an occurrent state of mind) and not necessary moved by representations (as cognitions). This kind of extension is existential and constitutes the subject as "being with" (Nancy 1996, Candiotto 2016b). As Trevarthen has claimed, the self-regulation is *alteroceptive*[8]. Affectivity is the more primitive way in which the subject is open to the others and constitutes itself as "being with", and the primordial intersubjectivity is the source for protocognition as affectivity.

5. Conclusion

Having discussed PAA and EET and having provided some very specific points of difference between them, I introduced my approach of extended affectivity as primary intersubjectivity (EACPI), arguing for the protocognitive value of affectivity. Moreover, I explained why the distributed cognition model is the more convenient to understand, at the same time, the collective and the subjective dimension of EA.
I think that this topic is both intriguing and tricky and that further investigations are required, especially regarding the elaborations of replies to the critics of the representationalists for whom, since cognition is representation, it cannot be performed within primary intersubjectivity. Although more work is required in this field of research, it seems fair to conclude that one reply may arise from an analysis of the embodied access to intersubjectivity as protocognitive and of the primordial empathic responsiveness.

REFERENCES

Candiotto, L. (2016a), "Emotions", In Peters, M. (ed.), *Encyclopedia of Educational Philosophy and Theory*, Springer, Singapore, http://link.springer.com/referenceworkentry/10.1007/978-981-287-532-7_311-1;
- (2016b), "Heidegger's ontic relatedness: *pros ti* and *Mitsein*", *Aurora, Revista de Filosofia*, Vol. 28, No. 43, pp. 313-328;
- (2015), "Aporetic State and Extended Emotions: the Shameful Recognition of Contradictions in the Socratic Elenchus", *Etica & Politica/Ethics & Politics*, Vol. XVII, No. 2, pp. 233-248;
Carter, J. A., Gordon, E. C., and Palermos, S. O. (2016), "Extended emotion", *Philosophical Psychology*, Vol. 29, No. 2, pp. 197–218;
Clark, A. (2016), *Surfing uncertainty. Prediction, Action, and Embodied Mind*, Oxford University Press, Oxford;
-(1997), *Being There: Putting Mind, Body, and World Together Again*, MIT Press Cambridge (MA).
Clark, A., and Chalmers, D. (1998), "The extended mind", *Analysis,* 58, pp. 10-23;
Colombetti, G. (2015), "Extending the extended mind: the case for extended affectivity", *Philosophical Studies*, Vol. 172, No. 5, pp. 1243-1263;

7 Fuchs and Koach (2014) have demonstrated the long-term effects of the embodied interaffectivity through the recognition of the "intercoporeal memory".

8 "Social animals have evolved ways of inter-acting, of moving in complementary or coordinated ways and sensing the power and purposes of one another's movements intersubjectively, so that, by cooperating in their motives, they can increase their individual and collective benefits and their adaptation in a sustaining ecology. They make their autonomic self-regulations of vital state apparent to one another so they can act emotionally in 'sympathetic' ways.[...] Their social life requires that they evaluate the purposes of one another by detecting intentions, interests and feelings from the energetic and self-regulating qualities of each other's movements, alteroceptively", Trevarthen 2011: 123.

- (2014), *The Feeling Body: Affective Science Meets the Enactive Mind*, MIT Press, Cambridge, MA/ London;
- (2013), "Some ideas for the integration of neurophenomenology and affective neuroscience", *Constructivist Foundations,* Vol. 8, No. 3, pp. 288-297;
Daly, A. (2014), "Primary Intersubjectivity: Empathy, Affective Reversibility, 'Self Affection' and the Primordial 'We'", *Topoi*, 33, pp. 227-241;
Deigh, J. (1994), "Cognitivism in the Theory of Emotions", *Ethics*, 104, pp. 824–54;
De Sousa, R. (2014), "Emotion", In Zalta, E. E. (ed.), *Stanford Encyclopedia of Philosophy*;
Freeman, W. (2000), "Emotion is essential to all intentional behavior", In Lewis, M. D., and Granic, I. (eds.), *Emotion, Development, and Self-Organisation: Dynamic Systems Approaches to Emotional Development*, Cambridge University Press, Cambridge, pp. 209-235;
Fuchs, T., and De Jaegher, H. (2009), "Enactive Intersubjectivity. Participatory sense-making and mutual incorporation", *Phenomenology and the Cognitive Sciences*, Vol. 8, No. 4, pp. 465-486;
Fuchs, T., and Koch, S. (2014), "Embodied Affectivity: on moving and being moved", *Frontiers in Psychology*, Vol. 5, No. 508, pp. 1-12;
Gallagher, S. (2013), "The socially extended mind", *Cognitive Systems Research,* 25–26, pp. 4–12;
- (2001), "The practice of mind. Theory, simulation, or primary interaction?", *Journal of Consciousness Studies,* 8, pp. 83-103;
Gilbert, M. (1989), *On Social Facts,* Princeton University Press, Princeton;
Goldie, P. (2007), "Seeing what is the kind of thing to do. Perception and emotion in morality", *Dialectica,* Vol. 61, No. 3, pp. 347–361;
James, W. (1950), *The Principles of Psychology*, Dover Publications, New York;
Krueger, J. (2014), "Varieties of extended emotions", *Phenomenology and the Cognitive Sciences*, Vol. 13, No. 4, pp. 533-555;
- (2013), "Ontogenesis of the socially extended mind", *Cognitive Systems Research*, 25-26, pp. 40-46;
Lewis, M. D. (2005), "Bridging emotion theory and neurobiology through dynamical systems modeling", *Behavioral and Brain Sciences,* 28, pp. 169-245;
Menary, R. (2006), "Attacking the bounds of cognition", *Philosophical Psychology*, 19, pp. 329–344;
Nancy, J. L. (2002), *La création du monde ou la mondialisation,* Galilée, Paris;
- (1996), *Être singulier pluriel,* Galilée, Paris;
Noë, A. (2004), *Action in Perception*, MIT Press, Cambridge, MA;
- (2010), *Out of our heads. Why you are not your brain,* MacMillan, Hill and Wang, London;
Palermos, S. O., and Pritchard D. (2013), "Extended Knowledge and Social Epistemology", *Social Epistemology Review and Reply Collective*, Vol. 2, No. 8, pp. 105-120;
Sauer, H. (2012), "Psychopaths and Filthy Desk. Are Emotions Necessary and Sufficient for Moral Judgments?", *Ethical Theory and Moral Practice*, Vol. 15, No. 1, pp. 95-115;
Slaby, J. (2014), "Emotions and the Extended Mind", In Salmela, M. and von Scheve, C. (eds.), *Collective Emotions,* Oxford University Press, Oxford;
Stephan, A., Walter, S., and Wilutzky, W. (2014), "Emotions beyond brain and body", *Philosophical Psychology*, Vol. 27, No. 1, pp. 65-81;
Thompson, E., (2007), *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind,* Harvard University Press, Cambridge, MA;
Trevarthen, C. (2011), "What Is It Like to Be a Person Who Knows Nothing? Defining the Active Intersubjective Mind of a New Born Human Being", *Infant and Child Development,* 20, pp. 119-135;
- (1979), "Communication and cooperation in early infancy: a description of primary intersubjectivity", In Bullowa, M. (ed.), *Before Speech. The beginning of interpersonal communication*, Cambridge University Press, Cambridge, pp. 321-348;
Varela, F. J., Thompson, E., and Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience,* MIT Press, Cambridge, MA.

ROBERTA PATALANO
Università Parthenope
roberta.patalano@uniparthenope.it

ON THE CONSTRUCTIVE ROLE OF CONFLICTING EMOTIONS: THE CASE OF EARLY MOTHER-CHILD INTERACTION AND ITS RELEVANCE FOR THE STUDY OF SOCIAL BEHAVIOUR

abstract

In this paper I seek to define a first, preliminary basis for a dialogue between philosophy and psychoanalysis on the topic of emotional conflict. As I will argue, the interaction between mother and child in the latter's first year of life represents a privileged vertex of observation on the positive effects that can be produced by coping with emotional ambivalence, both on the quality of the relationship and on the development of the child. Furthermore, tolerance for emotional conflict not only contributes to the development of the Self in the infant but it also favours the acquisition of prosocial attitudes, such as the capacity for concern, authenticity and creativity.

keywords

ambivalence, integration, authenticity, concern, morality

Phenomenology and Mind, n. 11 - 2016, pp. 242-250
DOI: 10.13128/Phe_Mi-20123
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In a well-known passage of the *Symposium,* Aristophanes talks about the mythological origin of love. Humans, he says, were cut in half by Zeus who wanted to punish them for their arrogance and their attempt to overthrow the gods. This punishment decreased their strength and made them more dependent, both on the gods and on each other. This original wound gave humans a permanent desire to recreate their initial unity and search for their lost half. This myth, according to Aristophanes, explains the origin of love, which can be essentially interpreted as a desire, in particular as the desire for Completion and Wholeness.

In a subsequent speech, Diotima interprets love as a basic human motivation, a vital force which generates the passions that animate human life. But - and through this argument she specifies Aristophanes's myth - love is a desire for the Good only. No one desires anything that is bad. In particular, "even if lovers seek their other half for the sake of wholeness, they only want to be a whole with the good part, while the parts that are bad will be avoided and kept away. Good things, in their turn, are not desired for their own simple goodness, but are sought as the means to achieve happiness which is humans' final end" (Naugle 2016).

Plato's idea of love as always directed towards what is good diverges significantly from the psychoanalytic view of love and hate as inextricable and omnipresent in our affective relationships. According to the latter, love has a dark, ambivalent side and is also directed to those parts of the Other that may hurt and frustrate us.

Since its inception, psychoanalysis has identified emotional ambivalence as a distinctive characteristic of our psychic functioning. A seminal development of the concept of emotional ambivalence was provided by Melanie Klein (1937, 1946, 1948, 1952). Afterwards, building upon Klein's theory, Winnicott (1945, 1949, 1960, 1963) focused on the ambivalence of maternal affect. In his article "Hate in the Counter-Transference" (1949), a famous paragraph describes at least eighteen reasons that a mother has to hate her child before the child learns to hate her. These include: "The baby is not her own (mental) conception", "The baby is not one of childhood play, father's child, brother's child etc." "The baby is ruthless, treats her as a scum, an unpaid servant, a slave" (Winnicott 1949: 72-73).

In this paper I argue that the emotional ambivalence theorised by Winnicott and Klein is not a symptom of dysfunctionality but has a constructive role and is functional for living and perceiving a life as authentic.

Furthermore, I suggest that the personal elaboration of ambivalence is not only significant for individual maturation but it also promotes, or inhibits, the emergence of a natural basis for collective morality.

How may the psychoanalytic perspective on love and hate be reconciled with Plato's view in the *Symposium*? I answer this question by focusing on the difference between the intelligibility of emotional conflict and intimacy with it. Although both are needed, they serve different functions, as I will explain.
In particular, intimacy with ambivalence is constitutive of a mature functioning modality of the psyche in which all components of reality can be perceived and accepted. Far from being only a personal achievement, such a capacity also has significant implications for society, as I will tentatively suggest.

2. Love and Hate in Melanie Klein's Contribution

Emotional ambivalence is the experience of having *simultaneously* both positive and negative feelings toward an object. Such an experience gives rise to an internal conflict in the subject, who feels love and hate at the same time, and toward the same person.
According to the English psychoanalyst Melanie Klein (1946, 1948), the coexistence of love and hate characterises human beings from their earliest days of life. Klein develops her argument through the analysis of the most primitive mental states that arise in children. According to her most famous theoretical contributions, the infant develops an integrated Self by going through the turbulent phases of the paranoid-schizoid position and the depressive position in the first year of life. The vicissitudes of these phases - how they are faced and dealt with in the relationship with, mostly, the mother - have an impact on the infant's maturation and on her chance to reach psychic integration.
Klein (1946) uses the term "paranoid-schizoid position" to describe the earliest months of an infant's life. Its most perspicuous evidence is the splitting of both the Self and the Object (the person to whom feelings are directed) into a "good part" who loves and a "bad part" who hates. The good and the bad part are not integrated but coexist in the Self and in the Other. The infant uses phantasies of splitting, projection and introjection in order to deal with fear and anxiety that arise from her conflicting feelings. The most intense conflict has to do with the coexistence of love and hate in the relationship with the mother. She uses splitting to manage this unbearable ambivalence. Once her own feelings have been separated into loving and hating ones, she projects them into separate parts of the mother. As a result, the mother is also split into a good mother, or breast, from which love and gratification can be received, and a bad one, that is felt to be frustrating and persecutory.
If development proceeds normally, splitting will be gradually given up, the mother who is hated and the mother who is loved will be "put together" as parts of the same person and, accordingly, the Self will be felt as an integrated and tolerable mixture of good and bad components. But this stage will take time to be consolidated.
In the paranoid-schizoid position, binary splitting is still dominant and, in parallel, omnipotence and idealisation colonise the mind of the infant. Bad experiences are omnipotently denied and projected outside the Self to reduce persecutory feelings, while good experiences are idealised as a protection against the fear of the persecuting breast[1].
The depressive position follows the paranoid-schizoid one. It begins in the second six months of life and is repeatedly refined throughout childhood and intermittently throughout life. In this phase, love and hate still coexist but love surmounts hate, thus promoting integration. At this stage of development, the baby realises and accepts that the hated mother and the loved mother are one and the same person. By re-composing the ideal and the persecuting

1 Both constitutional and environmental factors affect the course of the paranoid-schizoid position. Constitutional factors include the life and death instincts in the infant and the balance between them. The main environmental factor is the mother and the quality of the maternal care that the infant receives.

parts of the Object, she is enabled to re-compose love and hate in her feelings, thus developing a more integrated sense of Self.
As a fundamental consequence, the Self, the Object, and the world are perceived more richly and realistically. The Object is recognised as separate from one's own Self and as real, while idealisation and illusion of omnipotent control over it diminish.

3. Winnicott and the Ambivalence of Maternal Affect

Winnicott[2] built upon Klein's perspective by focusing on the mother-child interaction in the first year of life. The *good enough mother* that Winnicott (1945) describes is full of ambivalences about being a mother. When she takes care of her baby, she is selfless and generous, offering dedication and deep love, but she is also self-interested and prone to resentment.
A mother has many and varied reasons to hate her baby as Winnicott suggests in his article "Hate in the Counter-Transference" (1949: 72-73):

> The baby is not her own (mental) conception;
> The baby is not the one of childhood play, father's child, brother's child, etc.
> The baby is not magically produced;
> The baby is a danger to her body in pregnancy and at birth;
> The baby is an interference with her private life, a challenge to her preoccupations;
> To a greater or lesser extent, a mother feels that her own mother demands a baby so that her baby is produced to placate her mother;
> The baby hurts her nipples even by suckling, which is at first a chewing activity;
> The baby is ruthless, treats her as scum, an unpaid servant, a slave;
> She has to love him, excretions and all, at any rate until he has doubts about himself;
> He tries to hurt her, periodically bites her and all, we say, in love;
> He shows disillusionment about her;
> His excited love is cupboard love, so that having got what he wants, he throws her away like orange peel;
> The baby at first must dominate: he must be protected from coincidences, life must unfold at the baby's rate, and all this needs his mother's continuous and detailed study;
> At first he doesn't know at all what she does or what she sacrifices for him and especially he cannot allow for her hate;
> He is suspicious, refuses her good food, makes her doubt herself, but eats very well with his aunt;
> After an awful morning with him, she goes out and he smiles at a stranger who says, "Isn't he sweet?";
> If she fails him at the start, she knows he will pay her out and for ever;
> He excites her (sexually too) and frustrates her – she must not eat him or trade in sex with him.

As this quotation suggests, a mother has many comprehensible reasons to hate her own child. What seems to be important, however, is not hate as such but its acceptance. The process of coping with ambivalence[3] is indeed fundamental and requires an acknowledgement both of its

2 Donald Winnicott (1896-1971) was an English pediatrician and psychoanalyst who devoted much of his professional life to the study of mothers and babies. Through his concepts of the "good enough mother" and the "transitional object", he revolutionized psychoanalytic thought but also became a popular writer and contributor to the public debate on education and child care.
3 According to Winnicott (1945, 1949), such a process requires recognition of ambivalence as an unavoidable part of the mother-infant relationship. Only when this recognition takes place signification of ambivalence within the

existence and of its potentially positive effects. Let me consider the effects for the mother first. In my view, emotional ambivalence helps the mother in two main ways:

- The mother who is able to accept the conflictual nature of her feelings toward her baby without denying or minimising them will spontaneously resist the temptation to offer the baby an idealised image of herself, because she experiences the dark, ambivalent nature of her feelings and this makes her feel human, limited and imperfect. She will be authentic in the relationship with her baby.
- Such a mother is much more likely to gradually forget the "ideal baby" that she imagined during pregnancy and to start a relationship with the real one, the baby as she truly is, regardless of how different she may be from any idealised image. This will put her in touch with the authentic baby.

Let me now consider the infant[4]. In the first year of life, the infant's Self is in development and depends on the maternal Self to emerge and become stable. In this phase, the infant exists only because of maternal care, together with which it forms a unit (Winnicott 1945).
In particular, any infant starts with an inherited potential that can develop or vanish according to the quality of maternal - and then parental - care which is provided. Winnicott uses the expression "facilitating environment" to define a context in which the needs of the baby are sufficiently understood and met by her parents, thus guaranteeing the natural course of development:

> Infants come into *being* differently according to whether the conditions are favourable or unfavourable. At the same time conditions do not determine the infant's potential. This is inherited, and it is legitimate to study this inherited potential of the individual as a separate issue, *provided always that it is accepted that the inherited potential of an infant cannot become an infant unless linked to maternal care* (Winnicott 1960: 589).

Through satisfactory parental care, the baby starts to feel integrated, she becomes a "unit", and, most importantly, she learns to differentiate "me" and "not-me". This latter ability gradually leads her to see the mother as a separate object. Gradual, relative independence is then acquired through a sort of "mental detachment" from the mother which allows for differentiation into a separate self.
A mother who acknowledges and accepts the ambivalence of her feelings toward the baby helps her development in at least two ways:

1) The baby, by experiencing the love but also the hate coming from her mother, will learn to legitimate - and thus to maintain - both love and hate in herself as a natural and not a dangerous possibility, thus accepting their intermittence and inextricability as a constitutive part of the relationship. Thus the baby has the chance to develop her whole Self, without ignoring or excluding any part, in particular not minimising or denying the hating part.
2) Through contact with the integrated Self of her mother, the baby can develop the capacity to accept "the good" and "the bad" in the Other and to enter into relationship with both.

These achievements are essential for living and perceiving a life as authentic, so they strongly contribute to the well-being of the individual.
As Winnicott argues,

relationship becomes possible.

4 According to Winnicott, the terms "infant" or "baby" imply a child whose relationship and communication with the mother is mediated by maternal empathy and not yet by language.

> sentimentality is useless for parents, as it contains a denial of hate, and sentimentality in a mother is no good at all from the infant's point of view. It seems to me doubtful whether a human child as he develops is capable of tolerating the full extent of his own hate in a sentimental environment. He needs hate to hate (1949: 73).

In other words, the baby needs to be in contact with the full spectrum of human emotions to learn that they all deserve legitimacy and that they all need to be preserved in order to develop an authentic Self.

4. Capacity for concern and moral sense

The nexus between emotional ambivalence and guilt further clarifies the importance of emotional conflict and its constructive potential in the affective world of the infant. Emotional conflict does not only allow for the development of an integrated Self and for the capacity to be in relationship with an integrated Other, but it also represents the basis for the acquisition of other fundamental psychic capacities. For both Klein and Winnicott, by coming to accept the reality of her own aggressive drives, the infant gradually starts feeling guilty for her hate and responsible for the damage that the Object may have suffered because of her.
In a new-born infant, whose personality is still far from integrated, destructiveness and hate are unaffected by concern. Initially, the impulse is ruthless, unaware of the love and the hate that brings it about. The mother's capacity to survive, however, allows for the emergence in the baby of a gradual capacity to differentiate between reality and fantasy, between the real mother who is still there and the phantasmatic mother who has been attacked and destroyed. Once the fantasy has been recognised as different from reality, the baby can feel fully responsible for the fantasy in all its dimensions - both positive and hostile.
With the acknowledgment of one's own aggressive drives, the capacity to experience a sense of concern for the damaged object develops. Such a capacity triggers the intention to repair, a constructive motivation that stimulates creativity. Indeed, the urge to repair is a vital sentiment that gradually strengthens, especially in the presence of a good enough mother who supports the child and shows - with her affective presence - that she has survived the hate and the aggressive attacks that were present in their relationship.
The nexuses that Winnicott has made explicit among tolerance of ambivalence, integration, sense of guilt, and reparation are not only significant for individual maturation. They also have a considerable social impact, for they promote, or inhibit, the emergence of a natural basis for collective morality (Winnicott 1963, 1966).
In fact, in the young child, tolerance of ambivalence allows for the spontaneous structuring of a moral sense - that is, the capacity to recognise the damage to the Other and to feel guilty about it (Winnicott 1958a). Such "primitive moral sense" is not inculcated in the mind from outside, nor is the result of an imposed code of conduct. Instead, it springs naturally from the dynamic of human affectivity, given that ambivalence has been recognised and accepted, first by the mother and then by the child[5].
As Winnicott argues,

5 In Winnicott's words, "in health the infant can hold the guilt, and so with the help of a personal and live mother (who embodies a time factor) is able to discover his own personal urge to give and to construct and to mend. In this way much of the aggression is transformed into the social functions, and appears as such. In times of helplessness (as when no person can be found to accept a gift or to acknowledge effort to repair) this transformation breaks down, and aggression reappears. *Social activity cannot be satisfactory* except it be based on a feeling of *personal* guilt in respect of aggression" (1950: 207).
Moreover, in unfavourable circumstances that prevents the emergence of the basic capacity for feeling guilty, "the implanted moral code is necessary but the resultant socialization is unstable" (Winnicott 1958b/1965: 24-25).

> The study of the sense of guilt implies for the analyst a study of individual emotional growth. Ordinarily, guilt-feeling is thought of as something that results from religious or moral teaching. Here I shall attempt to study guilt-feeling, not as a thing to be inculcated, but as an aspect of the development of the human individual. (...) Those who hold the view that morality needs to be inculcated teach small children accordingly, and they forgo the pleasure of watching morality develop naturally in their children, who are thriving in a good setting that is provided in a personal and individual way (1958: 14).

5. Intelligibility of, *vs*. intimacy with, emotional conflicts

Psychoanalysis has shed light on emotional ambivalence as an unavoidable human characteristic which can potentially promote healthy psychic maturation in the individual and the development of prosocial attitudes, such as moral sense and creativity.
Interestingly, in his *Ethics,* Spinoza anticipates a fundamental issue of psychoanalysis: the possible coexistence of opposite feelings for the same object of love: "If we conceive that a thing, which is wont to affect us painfully, has any point of resemblance with another which is wont to affect us with an equally strong emotion of pleasure, we shall hate the first-named thing, and at the same time we shall love it" (*Ethics*, quoted by Greenspan 1980: 225).
As this quotation shows, Spinoza does not only allow for a form of ambivalence between love and hate that he calls "fluctuation, or vacillation", but he also suggests a nexus between ambivalent feelings that are experienced simultaneously and the basic emotions of pleasure and pain coming from the same person (Greenspan 1980). Such a nexus has a clear affinity with the Kleinian idea of simultaneous gratification and frustration that the infant receives from the mother, due to the human impossibility to satisfy the Id - the impulsive part of the infant Self - in all its requests. This impossibility and the subsequent experience of pain and pleasure from the same affective source generate in the infant the inextricable mix of love and hate that Spinoza also describes.
Notwithstanding this affinity and although philosophers, as Geenspan (1980: 224-225) argues, nowadays know enough about emotions as "to question the familiar ideal of 'philosophic detachment' from them", emotional ambivalence remains a controversial issue (but see Berlin 1958), especially for those who consider emotions as a kind of rational judgement.
An analysis of the philosophical debate on emotions is beyond the purposes of this paper. Instead, I would like to suggest how philosophy and psychoanalysis could integrate their lines of inquiry on emotions, and especially on conflicts between opposite feelings.
I would like to start from this idea: the focal question is not how to free us from conflicts – as this is not possible and, were it possible, it would not be useful; but rather, the question is how we can become free to represent the conflicts in our mind and to feel involved in them. Representation of and involvement with emotional ambivalence are the neuralgic points. While philosophy can enrich our intelligibility of emotional ambivalence by investigating the modalities through which it can be represented, psychoanalysis can guide us to recognise it as a constitutive part of ourselves, a part with which we are in intimate contact throughout life. Intelligibility and intimacy are different forms of knowledge and each of them has a specific, valuable potential.
The first is the result of a thought process which can widen the burdens of our comprehension, by eliciting the possible meaning that we attach to feelings and, also, by enriching our capacity to express them.
Intimacy, on the other hand, is a mental and bodily condition which implies emotional contact. It makes comprehension less abstract and deeply rooted in a personal, and also physical, experience which is coloured by one's own way of feeling and interpreting.
By integrating intelligibility and intimacy, I suggest that emotional conflicts deserve attention,

can be studied and fruitfully understood but that they also need to be recognised as a component of our identity, a part that plays a significant role in our development, for it can favour, or inhibit, well-being and social adaptation.
While intelligibility has to do with an intellectual process of comprehension and, as such, it presupposes the existence of a mature thinking mind, intimacy defines the affective ground for the development of such a mind. Moreover, it can affect the modality through which the mind is able to think, by influencing its capacity to "see" the world - and thus also one's object of inquiry - in more or less realistic terms and to enter into relationship with all its parts, those which are compelling as well as the obscure, those easy to grasp as well as the ambiguous, with none of them being excluded or denied.
Furthermore, both intelligibility and intimacy with ambivalence need an affective environment to develop.
Intelligibility needs intellectual exchange, comparison and integration among different and interdisciplinary vertices of observation (Creighton 1922, Urban 1929). Intimacy requires - and also fosters - a mind/body active interchange and, most of all, the affective bond with another person (Mendelsohn 1982, Zaltzman 2007). On the one hand, it has its origin in the existence of drives or impulses that cannot be completely satisfied, e.g. it has a physical, bodily root. On the other, it characterises the dynamic of primitive affective experiences: an infant cannot learn to maintain - and thus cannot benefit from - ambivalence in herself unless she has a mother who accepts hate and keeps it within their relationship.
I also believe that intimacy with love and hate and intelligibility of their functions can be seen as mutually reinforcing processes that generate knowledge, provided that they are both recognised and considered. On the contrary, it would be a loss to take one and leave the other. For intelligibility which is not subsumed by deep emotional awareness is bounded to the surface of things. Equally, intimacy which is not supported by verbalisation and reasoning cannot become conscious and be valued in its many potentialities.

6. Conclusions

The psychoanalytic study of early mother-infant relationships[6] gives us an image of love that is rather different from that which Plato offered us in the *Symposium*. Indeed, authentic love is not directed to the good only. Instead, it involves reciprocal acceptance of both positive and hostile feelings in both the Self and the Other.
The emotional ambivalence that arises in early mother-infant relationships does not vanish in childhood. Instead, it is continually experienced throughout life as an unavoidable and valuable component of our relationship with others. Moreover, it offers crucial contributions both to the individual, by favouring her development and the maturation of different psychic functions, and to society by contributing to the generation of prosocial forms of behaviour.
As far as the individual is concerned, the capacity to accept and maintain both love and hate for the Other in one's own conscious mental sphere favours integration and authenticity. It also sustains the passage to the depressive position and to a creative way out of it, which is based on concern for the Other and trust in reparation.
Pro-sociality, in its turn, is strengthened by emotional ambivalence in at least three main ways:

- First, the acceptance and personal development of ambivalence reduces the risk of acting it out, e.g. of an aggressive and uncontrolled attack on the Other;
- Secondly, it helps to mentally integrate reality in its multiple dimensions, thus affecting the capacity for realistic thinking;

6 The same argument may be applied to any primary care-giver who is in close and stable relationship with the infant.

- Thirdly, it is the basis from which moral sense springs naturally and sets its deepest roots, before being reinforced by culture and education.

REFERENCES

Creighton, J. E. (1922), "The Form of Philosophical Intelligibility", *The Journal of Philosophy*, 19, 10, pp. 253-261;
Greenspan, P. (1980), "A case of mixed feelings: Ambivalence and the logic of emotions", In Rorty, A. O. (ed.), *Explaining Emotions*, University of California Press, CA, pp. 223—250;
Klein, M. (1937), "A Contribution to the Psychogenesis of Manic–Depressive States", *International Journal of Psycho-analysis*, 16, pp.145-174;
Klein, M. (1946), "Notes on some schizoid mechanisms", *International Journal of Psychoanalysis*, 27, pp. 99-110;
Klein, M. (1948), "A contribution to the theory of anxiety and guilt", *International Journal of Psychoanalysis*, 29, pp. 113-123;
Klein, M. (1952), "Some Theoretical Conclusions Regarding the Emotional Life of the Infant", In Riviere, J. (ed.), by Melanie Klein et al., *Developments in Psycho-analysis,* Hogarth, London, pp 198-236;
Mendelsohn, R. (1982), "Intimacy in Psychoanalysis", In Fisher, M., and Stricker, G. (eds.), *Intimacy*, Plenum Press, New York;
Naugle, D., "The Platonic Concept of Love: The Symposium", http://www3.dbu.edu/naugle/pdf/platonic_love.pdf, retrieved on January 2016.
Urban, W. M. (1929), *The Intelligible World: Metaphysics and Value*, George Allen And Unwin Limited, London; reprinted by Routledge in 2014;
Winnicott, D.W. (1945), "Primitive emotional development", *International Journal of Psychoanalysis*, 26, pp. 137-143;
Winnicott, D.W. (1949), "Hate in the counter-transference", *International Journal of Psychoanalysis*, 30, pp. 69-74;
Winnicott, D. W. (1950), "Aggression in relation to Emotional Development", *Mental Hygiene*, 34(3), pp. 353-363;
Winnicott, D. W. (1958a), "The Antisocial Tendency", In Winnicott, D. W., *Collected Papers. Through Paediatrics to Psycho-Analysis*, Tavistock Publications, London, pp. 306-315;
Winnicott, D. W. (1958b), "Psychoanalysis and the Sense of Guilt", In Winnicott, D. W., *The Maturational Processes and the Facilitating Environment*, Karnac, London, 1965, pp. 14-28;
Winnicott, D. W. (1960), "The Theory of the Parent-Infant Relationship", *International Journal of Psychoanalysis*, 41, pp. 585-595;
Winnicott, D.W. (1963), "The development of the capacity for concern", *Bulletin of the Menninger Clinic*, 27 (4), pp. 167-176;
Winnicott, D.W. (1966), "The absence of a sense of guilt", In Winnicott, D. W., *Deprivation and Delinquency*, Tavistock Publications, London, 1984, pp. 91-96;
Zaltzman N. (2007), *L'esprit du mal*, Éditions de l'Olivier, Paris, trad. it. *Lo spirito del male*, Borla, Roma, 2011.

50,00 €